UNIVERSITY OF NORTH TEXAS

BENCHMARKS

The University of North Texas Computing Center Newsletter

VOLUME 13 NUMBER 6
July/August 1992

Statistical Computing At UNT

By Panu Sittiwong, Research and Statistical Support Manager (panu@unt.edu)

One of the main functions of Academic Computing Services is to provide statistical and research support to faculty and students at UNT. To help us achieve this goal, we have various statistical packages,running under various operating systems, available for a variety of computer hardware configurations. This article will discuss the statistical packages which are available at UNT and attempt to help you choose the appropriate software and platform to fit your particular needs.

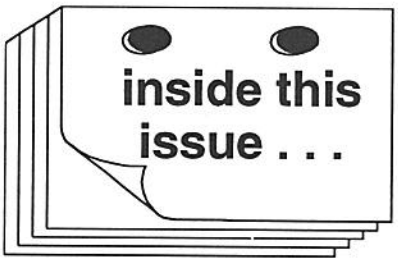
Host System Statistical Packages

Academic host system computers used for research statistical purposes at UNT are an IBM 3081 compatible mainframe computer (an HDS 8083) and a Solbourne superminicomputer. Currently, there are three major operating systems available to academic users on the IBM 3081: VM/CMS/XA, OS/MVS, and MUSIC/SP. The MUSIC/SP platform, however, is in the process of being phased out (beginning September 1, 1992). The Solbourne computer is running the UNIX operating system. The table on page 3 shows statistical packages available on the three platforms.

Microcomputer Statistical Packages

There are a large number of microcomputer packages available at UNT. In fact, each department has adopted particular packages that are appropriate to its needs and discipline. For example, the College of Business adopted Minitab PC for its instructional use. The Political Science Department, on the other hand, has a site license for RAT PC to be used in its classes.

Academic Computing Services maintains a site license for SAS/PC and SPSS/PC+ and SPSS for Macintosh. These three packages are available free of charge to all full-time faculty and staff at UNT. Teaching fellows and teaching assistants may request to have the software installed on a PC that belongs to the University. These packages are also available to students in all of the General Access Computer Labs on campus.



❑ RESEARCH
SAS Version 6.07 4
Using Secondary Data Sources 8
Researching the National Archives .10
CRSP Data Information 10
Formatting Theses and Dissertations in WordPerfect 12
The Network Connection 12
List of the Month 13
FEDIX 14
Using the SPSS Macro Facility 14
Violation of Assumptions 16
❑ GENERAL INFORMATION
Assistant to Associate V.P. for Computing Hired 18
Staff Activities 18
Presidential E-mail 18
IRC News 18
1992 Summer Short Courses 19
Calls for Papers 20
Accessing the UNT Library Catalog 20
❑ MICROCOMPUTERS
MICRO-TIPS 21
Virus Update 26
WPUG Summer Schedule 26
❑ VAX/UNIX SYSTEMS
VAXCLUSTER Usage Statistics 27
Sol: A UNIX System for Research ... 27
Solbourne Usage Statistics 28
The Coalition for Networked Information 29
❑ MAINFRAME TECHNICAL SERVICES
Performance Statistics 30
Disk Backup Schedules 31

☆ UNT COMPUTING CENTER ORGANIZATION AND FACILITIES ☆

The UNT Computing Center is located in the Information Sciences Building (ISB), Room 119. Phone: (817) 565-2324 (TDD 1-800-RELAY-TX), unless otherwise noted. It is divided into the following areas:

☐ **Academic Computing Services:**
- Documentation Services
- ISB 110 General Access Lab (817) 565-3048
- Mainframe User Services
- Statistical Services
- VAX/UNIX Systems (817) 565-4161

☐ **Network & Microcomputer Services** (817) 565-2316:
- Data Communications
- Microcomputer Application Support
- Network Systems Support

☐ **Administrative Computing:**
- Admissions Data Systems
- Database/Central Programming Support
- General Data Systems
- NT/TCOM Fiscal Data Systems
- NT/TCOM Payroll/Personnel
- Student Records
- Student Services
- Voice Response Applications

☐ **Mainframe Technical Services:**
- IBM Operating Systems Software Support
- Computer Operations

☎ CONNECTING TO UNT COMPUTERS ☎

Phone numbers for accessing UNT computing systems:

300-2400 BAUD: (817) 565-3300
300/1200 BAUD: (817) 565-3499
300-9600 BAUD: (817) 565-3461
300-2400 BAUD: D/FW METRO 792-4140

Area code 214 must dial 817 before the METRO #. ***Note:* Dialing 1 before the area code will result in a long distance charge.**

Set Data Bits to **7**, Parity to **S**, and Stop Bits to **1**. When dialing in, the autobaud feature requires you to hit the <RETURN> key repeatedly after the connection is made so that the receiving modem can determine the baud rate. When you see the prompt (# for local numbers, UNTModems> for the metro lines) you can enter one of the following commands to connect with the system of your choice.

SYSTEM	*SYTEK/DENTON LINES (#)*	*METRO LINES (UNTMODEMS>)*	*INTERNET (CUTCP, NCSA)*
ACS Host Systems			
MUSIC/SP (line editing & PCWS)	**CALL 8040**	**N/A**	**N/A**
Academic Mainframe (MUSIC, CMS, Academic COM-PLETE)	**CALL 3270**	**CONNECT VM3270**	**tn3270 vm.acs.unt.edu —OR— telnet vm3270.acs.unt.edu**
VAX (VMS)	**CALL DEC**	**CONNECT DEC**	**telnet vaxb.acs.unt.edu**
Solbourne (UNIX)	**CALL 900**	**CONNECT SOL**	**telnet sol.acs.unt.edu**
Departmental Systems			
Computer Sciences **Sequent** (Ponder)	**CALL 780**	**CONNECT PONDER**	**telnet ponder.csci.unt.edu**
UNT Libraries' on-line card catalog	**CALL 3000**	**CONNECT LIBRARY**	**telnet library.unt.edu**

To exit from the local phone lines, press <ESCAPE><RETURN>, and type DONE (at the # prompt), then press <RETURN><RETURN>. To exit from the metro lines, press <CTRL-SHIFT-6>, then type DISCONNECT (at the UNTMODEMS> prompt), then press <RETURN>. Exiting from telnet and TN3270 is dependent upon the package. CUTCP uses <ALT-X>.

HOURS FOR UNIVERSITY OF NORTH TEXAS COMPUTER ACCESS AREAS : Summer 1992

Day of Week	*Willis*	*BA*	*ISB 110*	*Chilton, Matthews*	*GAB*	*Terrill, Wooten*	*ISB 205C*	*Lab Locations*
Monday - Thursday	Open 24 hrs.	8 a.m. - 11:45 p.m.	7:30 a.m. - 10 p.m.	8 a.m. - 10 p.m. **CH 255** - 6 p.m.	8 a.m.- Midnight	8 a.m. - 6 p.m.	Noon - 10 p.m.	✪ **BA:** 330, 332 ✪ **Chilton :** 255, 116 [Adaptive Lab] ✪ **GAB:** 330, 550 ✪ **ISB:** 110, 205C — grad. students only ✪ **Matthews:** 309 ✪ **Terrill** 247 ✪ **Willis:** 134 ✪ **Wooten:** 120
Friday	Open 24 hrs.	8 a.m. - 8 p.m.	7:30 a.m. - 6 p.m.	8 a.m. - 5 p.m.	8 a.m.- 3 p.m.	8 a.m. - 3 p.m.	1 p.m.- 5 p.m.	
Saturday	Open 24 hrs.	8 a.m. - 8 p.m.	9 a.m. - 6 p.m.	10 a.m. - 5 p.m.	Closed	Closed	10 a.m. - 5 p.m.	
Sunday	Open 24 hrs.	Noon - 11:45 p.m.	1 p.m. - 10 p.m.	1 p.m. - 10 p.m. **Chilton 255 closed**	2 p.m.- Midnight	Closed	1 p.m. - 5 p.m.	

This issue of **Benchmarks** *was produced by the Documentation Services section of Academic Computing using Xerox Ventura Publisher on a 386SX clone and printed on an HP LaserJet IIISi.* *Unless otherwise noted, articles or information may be reproduced for nonprofit purposes provided the publication and issue are fully acknowledged.*

Host System Statistical Software at UNT

	UNIX	VM/CMS	OS/MVS
SAS Version 6.07	All modules	Base, Stat, ETS, OR, Graphs, IML	Base, Stat
SAS Version 5.18	N/A	Base, Stat, Graph	Base, Stat
SPSS Version 4.1	N/a	Basic, Stat, Lisrel 7	Basic, Stat, Lisrel 7
Shazam Version 6.2	N/A	All modules	All modules
BMDP 90	N/A	N/A	All modules

If you need to have any of these packages installed on your microcomputer, you may contact one of the ACS statistical consultants at (817) 565-2324 or you can mail your request to PANU on either the CC1 file server (**CC1:PANU** for WordPerfect Office Mail or **PANU@cc1.unt.edu** on Pegasus mail) or **PANU@untvm1** on VM/CMS or **PANU@untmusic** MUSIC/SP mail. The software will take about 12MB of hard disk space. You will need to check your hard disk space to verify that you have enough space for the software since our installer program doesn't allow for the removal of any files from your disk.

Currently, ACS does not support any MS-Windows based or OS/2-based statistical software. SAS Institute Inc. is currently testing the SAS system for Windows and anticipates having it available before the end of the year. SPSS Inc. has recently released SPSS for Windows. The product, however, only includes Base and Basic Stat. The company anticipates having the Adv. Stat. module available in the third quarter of this year. ACS is currently evaluating the SPSS for Windows product. If we acquire the product, it will probably be available during the fall semester.

Platforms for Statistical Computing

Most of the time, the platform on which you run your statistical packages is selected based on previous experience. This rationale, in some cases however, is in contradiction to changes taking place in computing. For example, MUSIC/SP will be deinstalled from the academic mainframe. This change will require all previous MUSIC/SP users to convert their mainframe work to either VM/CMS, UNIX, or a microcomputer application. The following section will present some of the advantages and disadvantages of the statistical package platforms which are currently available or will be available at UNT in the near future.

General Guidelines for Choosing a Statistical Computing Platform

SAS users at UNT have an advantage over users of some other statistical packages since SAS is available on all three platforms (see the table "Host System Statistical Software at UNT" above). SPSS users can select between SPSS on CMS, SPSS on OS/MVS, SPSS/PC+ or SPSS for the Macintosh.

Microcomputers (most of which are 286 or above machines) are now widely available on campus. Most researchers now have microcomputers in their offices. They provide a suitable platform for statistical computing as compared to the mainframe host system. This does not mean that you will no longer need to use the mainframe for your statistical computing, however. Both platforms provide different advantages and disadvantages to the user. This section will try to point out some of those advantages and disadvantages in order to help you choose the proper platform for your statistical research work.

The Microcomputer Platform

There are several options available to you when you select a microcomputer for your statistical computing platform. You may choose to run the software directly under the MS-DOS operating system, you may want to invest in MS-Windows, or you may decide on the Macintosh.

Both SAS and SPSS for microcomputers provide several advantages to users. These include:

1. SPSS/PC+ provides a Menu system which you can use to compose SPSS/PC+ programs. This helps to minimize syntax errors for new users. With this system, it is possible for a novice user to compose a SPSS/PC+ program without opening a manual.
2. Both SAS/PC and SPSS/PC+ provide excellent on-line help systems. All commands and syntax can be acquired from the help command.
3. Both systems have procedures which will allow you to incorporate the output into some other microcomputer program. For example, a Tables Procedure in SPSS/PC+ or PROC Tabulate in SAS/PC are capable of producing a customized table for your data. The output can be directly included in WordPerfect.
4. Graphic output from both packages can be directly incorporated into WordPerfect or DrawPerfect.

A major disadvantage of the microcomputer platform is its limited mem-

ory utilization. Under the MS-DOS operating system, SPSS/PC+ can only use up to 64K of Expanded Memory in addition to the conventional 640K RAM. This results in limiting the number of variables which can be used with SPSS/PC+. Currently, SPSS/PC+ has a limit of 500 variables. In addition, some procedures that require a large work space may not be able to be performed with SPSS/PC+. SAS/PC, on the other hand, can use up to 2M of Expanded Memory in addition to the conventional 640K RAM. This makes it a more flexible system. Also, if you are using data stored on tape (e.g., data from ICPSR, CRSP, or COMPUSTAT), you will be required to subset the dataset before it can be used with SAS/PC or SPSS/PC+.

The Windows operating environment can be used to overcome limited memory utilization. Both SAS and SPSS for Windows will be capable of using up to 16M of memory. This gives both packages capabilities similar to their mainframe counterparts, but with a Graphical User Interface (GUI). Hence, if you have a PC that is based on the 286 (or above) microprocessor, the Windows-based software products can provide an excellent platform for you. If you plan to upgrade your microcomputer to take advantage of the statistical packages under Windows, however, you will need to keep the following recommendations in mind.

- Plan to install a minimum of 6M of memory on your microcomputer. Both SAS and SPSS will run with less memory but with a performance degradation.
- Plan to install the math coprocessor (80x87) chip on your PC.
- If you plan to have the software installed on your hard disk, you will need at least an 80M hard disk.
- Plan to have an EGA or better Display.
- Plan to have a mouse installed on your PC.

Mainframe Platforms[1]

Obviously, the main advantage of running mainframe statistical software is the ability to utilize a large database in your research. Secondly, if you don't have a 286 or above microcomputer, the mainframe will provide a more powerful platform for you.

As mentioned earlier, statistical software is available either under OS/MVS or VM/CMS. In some cases, you may not have any choice but to run your statistical analysis on the mainframe under one or another operating system. Where a choice is available, we recommend that you run your analysis under the VM/CMS operating system. There are several advantages for running statistical software on VM/CMS. These advantages include:

- All of the above software can be run interactively on VM/CMS. In this environment, you will have a software interface similar to that of a microcomputer.
- Since the CMS operating system is interactive, all jobs submitted will be executed immediately.
- SAS users have the option to use both SAS/PC and SAS/CMS simultaneously. It is possible to submit a SAS/PC program to be executed on VM/CMS.
- With the implementation of TCP/IP on VM/CMS, users have the ability to move data between PCs and the mainframe quickly. Therefore, a broader range of options can be selected.

In conclusion, Academic Computing Services is committed to providing services in support of instructional and research activities at UNT. In order to achieve this goal, we provide a variety of services and software to our user community. This article was written in order to provide you with information to help you select the appropriate software and platform for your research purposes.

[1] This article does not discuss the UNIX platform since there is a limited amount of statistical software available on the Solbourne. It is a good platform if you are using SAS or IMSL subroutines which are computation intensive.

References

Cates, Mark W. "Taking Advantage of the Power and Flexibility of the SAS System," Paper presented at SUGI 17, Honolulu, Hawaii, May 10-12, 1992.

Sims, John and Dave Brumitt "Desktop Strategies: Should I Go to OS/2, Windows, or UNIX", Paper presented at SUGI 17, Honolulu, Hawaii, May 10-12, 1992. ■

SAS Version 6.07 for VM/CMS and OS/MVS

By Panu Sittiwong, Research and Statistical Support Manager (panu@unt.edu) and Phanit Laosirirat, Research and Statistical Consultant (phanit@unt.edu)

This article is the second in a series about the new SAS system at UNT.[1] There are several new features in this version of SAS. SAS version 6 is the first implementation of the SAS system which emphasizes compatibility across hardware platforms. To enhance the portability of the software, SAS Institute decided to rewrite the SAS system based on the C language. The first implementation of this version occurred in 1984 when SAS introduced SAS/PC. Besides portability, the new SAS provides several features which weren't available in version 5.18. This article will present the major enhanced features. It will not

[1] The first article appeared in the June 1992 issue of *Benchmarks* (Vol. 13 No. 5 p. 17).

cover all the changes and enhancements but rather will point out some major changes. The complete explanation of all the changes and enhancements can be found in the most recent edition of the appropriate SAS manual. (See the end of this article for a complete list of SAS manuals appropriate for UNT.) For example, a summary of the enhancements of SAS/BASE can be found on pages xxiii-xxix of the *SAS Language: Reference, Version 6 First Edition.*

The enhancements to SAS found in version 6.07 can be categorized as enhancements in general, enhancements in the SAS Data Step, and enhancements in SAS PROCs.

Enhancements in General

Version 6 is a new design of the SAS system, with the implementation of "Multiple Engine Architecture." With this system, you are able to access a variety of different types of data files as if they are SAS data sets. For example, SAS will directly recognize SPSS-X system files and SPSS EXPORT files as if they were SAS datasets. This will allow you to directly analyze SPSS datasets with SAS PROCs without prior conversion.

The Multiple Engine Architecture is a two way street. It also allows you to use a SAS dataset in the format of other software system files.

SAS 6.07 under VM/CMS is the first large-scale interactive SAS product available at UNT. Although SAS previously had this available, it was not implemented here. Interactive SAS allows you to start using SAS effectively with minimal CMS knowledge and experience. SAS/PC users, in particular, will have the same user interface. SAS/CMS and SAS/PC are almost identical in their Display Manager System.

By having SAS available on CMS, it is possible for those who have SAS/PC installed on their microcomputers to run SAS/PC and SAS/CMS simultaneously. This allows you to analyze data stored on your PC using procedures on CMS and bring the output back to the PC. Academic Computing Services is currently working on a script file which will establish the connection between SAS/PC and SAS/CMS for you. When it is finished, we will place an announcement in *Benchmarks* and on all the host system news.

Another general enhancement is the ability to compile SAS Data Steps. In version 6.07, you can compile the data step programs and execute them at another time.

This is useful when you are reading several datasets in the same way at different times. For example, you can compile a data step which will read data from several waves of your survey. Instead of having to recompile the data step for each set of data, you can compile and save the data step and call it to read the data each time you need it. The following SAS program illustrates this procedure.

```
Title 'Sample SAS Program for Compiling a SAS Data Step';

LIBNAME stored 'A';

DATA Survey;
        INFILE datain;
        INPUT (V1-V100)(1.0);
        etc ... ;
```

The above program will compile the sample data step and save it in the library called **stored** on your CMS minidisk **A** with the filename of **Survey** and filetype of **stored.** The following SAS program will call the stored program and perform analyses on the data set called WAVE1 DATA A.

```
Title 'Sample SAS Program Using a Compiled SAS Data Step';

FILENAME datain 'wave1 data a';
LIBNAME stored 'A';

DATA PGM=stored.survey;
PROC PRINT;
etc ... ;
```

Another new feature of SAS 6.07 is the ability to subset SAS data sets in both the Data Step and the Proc Step. SAS introduced a new statement called **WHERE** which can be used in both the Data Step and the Proc Step. The following example shows how **WHERE** can be used in the Proc Step.

```
Title 'Sample SAS Program With WHERE Statement';

PROC PRINT DATA=survey;
LIBNAME stored 'A';

DATA PGM=stored.survey;
        VAR v1-v50;
        WHERE lastname LIKE 'Mc%' AND v6='Male';
RUN;
```

When this program is executed, SAS will only print those observations whose variable *lastname* starts with Mc (the % is used in this context as a wild card) and are male. This command can greatly enhance the efficiency of SAS processing time.

One change that has some negative impact is that this version of SAS will no longer support SUGI Supplemental Library PROCs. If you currently use any of those PROCs it will be necessary for you to find a comparable SAS-supplied procedure to do your analysis.

Data Step Enhancements

The majority of changes and enhancements in SAS 6.07 are in the Data Step. These improvements include:

- easier accessing of external data files and SAS data sets.
- introduction of new statements to simplify SAS programming.
- improved ARRAY processing.
- ability to create a virtual data set.

In previous versions, external data files and SAS system files are referred to by operating system commands, i.e., using the DD statement in OS/MVS or FILEDEF command in VM/CMS. In version 6.07, these files can be referred to by using the FILENAME, LIBNAME and INFILE statements. The FILENAME command assigns a nickname to a CMS file. The LIBNAME defines the minidisk on which SAS system data sets are stored, and INFILE is used to fetch data into the SAS program. The following program shows how these three commands are used.

```
Title 'Sample SAS Program';

FILENAME datain 'SURVEY DATA A';
LIBNAME sasfile 'A';

DATA sasfile.survey;
        INFILE datain;
        INPUT etc ... ;
RUN;
```

The FILENAME command is used to assign the name **datain** for the CMS file called SURVEY DATA A. LIBNAME assigned a SAS data library called **sasfile** which is located on CMS minidisk A. When the program is executed, SAS will retrieve the raw data from the SURVEY DATA A file. Since the DATA statement specified a two level name, SAS will create a permanent SAS system data set and store it on the CMS minidisk specified in the LIBNAME command. In the example, the permanent data set will be stored in the CMS file called **SURVEY SASFILE A**.

The SAS data set has been restructured to add database-style features such as indexing and data compression. Indexing can be either a regular index (based on one variable) or a composite index (based on more than one variable). Indexing of the variable provides two major benefits. First, indexing provides faster access to a small subset of observations represented in the index. Second, if a BY statement is used in the PROC Step, PROC SORT is not required. This will help speed up processing and save CPU time.

SAS version 6 data sets can also be compressed. A compressed data set will require less space to store the data. This feature is useful when you work with large data sets containing hundreds of data items or consisting of several thousand observations. Compressed data sets, however, have some trade-offs associated with them. For example, observations in the compressed data set can't be addressed by observation number. This will cause more CPU time to to be used to access the dataset.

Array processing has also been improved. Now you can:

- refer to an element in an array.
- assign an initial value to an element.
- create temporary array elements.
- create multidimensional arrays.

New Procedures

SAS Inc. made changes and enhancement to several PROCs in version 6. Printed output from many procedures has a new format, so you may have to spend a little more time interpreting your results the first time.

There are three new statistical procedures added to **STAT** product. They are:

❶ PROC LOGISTIC — performs Logistic Regression for the model with dichotomous or ordinal dependent variables.

❷ PROC TSCSREG — performs a regression analysis for time-series cross sectional data.

❸ PROC CALIS — can be used for analysis of covariate structures, linear structure equations with latent variables and path analysis.

Two utility procedures were added to the **BASE** system.

❶ PROC SQL — implementation of the SQL language in the SAS system. This procedure can be used to manipulate SAS data sets and generate reports.

❷ PROC REPORT — allows you to create a report template for SAS data sets. It is intended to replace report writing using the FILE and PUT statements.

A Matter of Conversion

SAS version 6.07 is currently available, on a test basis, on VM/CMS and OS/MVS. If you want to use SAS on CMS and you have been a MUSIC/SP

user, you will need to move your programs and data from MUSIC/SP to CMS. If your User-ID was not authorized for CMS, you can request that it be updated. Come by the Computing Center (ISB 119) and pick up the User-ID Update form and have it processed.

You can send SAS programs and raw data files to CMS with the MUSIC sendfile command. It has the following format:

```
SENDFILE filename TO idnn AT UNTVM1
```

Where **filename** is the name of MUSIC file you want to send, and **idnn** is your CMS User-ID.

The file will be stored in your CMS reader. To retrieve the file from your reader:

❶ Issue the CMS command RDRLIST to view your reader files.

❷ You can receive the file, i.e., move it to your CMS library, by entering the command **RECEIVE / fn ft fm** on the command area in front of the file name. **fn** is the CMS file name and **ft** is the CMS file type. You should give your SAS programs the file type **SAS**. **fm** is the filemode, it is usually A. You may have to type this command over the name in the rdrlist display.

You then can use XEDIT on CMS to remove all the JCL command lines, and make all the necessary changes before you submit the program to SAS.

New Documentation

SAS also reorganized the documentation for Release 6. It is possible to use Version 5 documentation, but you will need the new documentation if you plan to take full advantage of the new system. The following table provides a quick guide to SAS Version 6.07 manuals relevant to our installation here at UNT, along with their Version 5 equivalents. Please note that all of these documents are available in the ISB-110 Lab.

Version 5 Manuals	Version 6 Manuals
SAS User's Guide: Basics, Version 5 Edition.	*SAS Language: Reference, Version 6 First Edition* — This manual explains all the SAS System and Commands, including the use of SAS Display Manager.
	SAS Procedure Guide, Third Edition — contains syntax for all the SAS/BASE PROCs.
	SAS Language and Procedures: Version 6, First Edition — a good introductory book. It demonstrates several useful SAS techniques.
	SAS Companion for the CMS Environment, Version 6, First Edition — contains all information concerning using SAS under CMS.
	SAS Guide to Macro Processing, Version 6, Second Edition — demonstrates the use of SAS Macros.
	SAS Guide to the SQL Procedure: Usage and Reference, Version 6, First Edition — explains the implementation of SQL in the SAS system and contains all the documentation on PROC SQL.
SAS User's Guide: Statistics, Version 5 Edition	*SAS/STAT User's Guide, Version 6, Fourth Edition, Vols. 1 and 2* — contains syntax for all statistical procedures.
SAS/GRAPH User's Guide, Version 5 Edition	*SAS/GRAPH Software Version 6, Fourth Edition, Vols. 1 and 2* — contains syntax for all SAS/GRAPH procedures.
SAS/ETS User's Guide, Version 5 Edition	*SAS/ETS User's Guide, Version 6, First Edition* — contains syntax for all procedures in the Econometric and Time Series module.
SAS/OR User's Guide, Version 5 Edition	*SAS/OR User's Guide, Version 6, First Edition* — contains syntax for all procedures in the Operations Research module.
SAS/IML User's Guide, Version 5 Edition	*SAS/IML User's Guide, Version 6, First Edition* — contains all statements and samples of the IML product.

All manuals are available for purchase at the University Store.

Indexing a SAS dataset

Available since Release 6.06, SAS Index enables the SAS system to locate observations in a SAS data set more quickly and efficiently. There are two obvious benefits one can gain from indexing a SAS data set:

❶ Fast access to a subset of observations, especially when selecting observations by using the WHERE command in the SAS data step. For example, the SAS command **WHERE (YEAR BETWEEN 85 AND 90) AND (CUSIP=123456);** will select those observations that have CUSIP equal to 123456 and year values between 1985 to 1986. The processing time is reduced dramatically if the SAS data set is indexed by CUSIP and YEAR.

❷ Quick processing time when used with a BY statement. This includes using the BY statement in the SAS procedures and the SAS data step. For example, the following commands will merge two indexed SAS data sets into a new data set faster and without first using the SORT procedure.

```
DATA NEW;
MERGE ONE TWO;
BY CUSIP;
RUN;
```

Types of Indexes

Two or more variables can be used to index a SAS data set. The variables used to index a SAS data set are called *key variables*. A key variable can be a numeral or a character. SAS indexes are classified as simple or composite, according to the number of key variables making up the index.

Simple Index

A SAS data set is classified as a *Simple Index* when the SAS system locates observations by the values of one key variable. In a simple index, SAS gives the index name the same name as the key variable. You can create more than one simple index for a SAS data set.

There are two methods to create a simple index for a SAS data set in Release 6.07.

❶ **Indexing an existing SAS data set.** The following example shows the PROC DATASETS statements you can use to create two simple indexes.

```
PROC DATASETS LIBRARY=COMPSTAT;
     MODIFY INDUSTRY;
     INDEX CREATE CUSIP;
     INDEX CREATE CONAME;
     CONTENTS DATA=INDUSTRY;
RUN;
```

The CONTENTS command prints the contents information for the modified SAS data set, COMPSTAT.INDUSTRY.

❷ **Indexing a newly created SAS data set.** The following example uses the DATA command to create two simple indexes named CUSIP and CONAME.

```
DATA COMPSTAT.INDUSTRY(INDEX=(CUSIP CONAME));
```

Please see SAS 6.07 on page 9.

Using Secondary Data Sources for Your Thesis or Dissertation

By Panu Sittiwong, Research and Statistical Support Manager (panu@unt.edu)

When preparing your prospectus for your thesis or dissertation, you may want to keep in mind some publicly available secondary data sources. You may be able to use these data as suppliments to your primary data source or as the main data source for your research.

Academic Computing Services has a great deal of machine readable data available to all students and faculty members. These data are acquired from several sources including the Inter-university Consortium for Political and Social Research (ICPSR), Department of Labor, Center for Research in Security Prices (CRSP), Standard and Poor's (COMPUSTAT), etc.

- **CRSP Data:** Once a year, UNT receives data from the Center for Research in Security Prices. These data include:
  - Monthly NYSE and AMEX Returns and Master file.
  - Daily NYSE and AMEX Returns file.
  - Daily NASDAQ Returns and Master file.
  - Daily, Monthly, Quarterly, and Annual Markets Indices file.
  - National Market System Securities file.
- **COMPUSTAT II Data:** UNT also receives, once a year, updated COMPUSTAT II data from Standard and Poor's. The current holdings include:
  - Annual Primary, Supplementary, and Tertiary (Industrial and Research) file.
  - Annual Over the Counter file.
  - Annual Bank file.

- Annual Price, Dividends and Earnings file.

In addition, we are in the process of acquiring some quarterly data.

- **ICPSR Data:** Data from the Inter-university Consortium for Political and Social Research constitutes the majority of data archived at UNT. Currently, there are more than 200 data titles locally available. As a member of the Consortium, UNT students and faculty members can request any data made available by ICPSR. Data available from ICPSR cover a wide range of subjects and disciplines including:
  - Public Opinion Surveys
  - Election Studies —both the US and foreign countries
  - Congressional Roll Call
  - General Social Surveys
  - Health Interview Surveys
  - Consumer Expenditure Surveys
  - Government Finance
  - World Economic Indicators
  - Population Surveys
  - Census for the U.S. and foreign countries
  - EURO-BAROMETER

You can search and locate ICPSR data available at UNT by issuing the command **FINDICPSR** from either the MUSIC/SP or VM/CMS prompt. If the data that you need is not available locally, you can request the data by contacting me at 565-2324, or by sending E- mail to either PANU on the CMS or AC09 on MUSIC. It may take up to 4 weeks before the data can be accessed when it is ordered from ICPSR.

Data from Other Sources

In addition to the above data sources, you can acquire data from other sources. If you plan to acquire data from another source, contact ACS concerning the method and medium of data storage before you order the data. ■

SAS 6.07 continued from page 8

Composite Index

A composite index refers to a SAS data set that has an index comprised of two or more key variables. The variables used in the composite index can be numerals, characters, or alphanumeric. To create a composite index of a SAS data set, you must specify a unique name for the index. The index name cannot be any name that has been used in the data set before. This includes the variable names. There are two methods to create a SAS data set with a composite index.

❶ **Indexing an existing SAS data set.** The following example shows the PROC DATASETS statements that create a composite index named COMPANY for the SAS data set COMPSTAT.INDUSTRY. The SAS data set contains two variables, INDUSID and CUSIP, that are used as key variables.

```
PROC DATASETS LIBRARY=COMPSTAT;
        MODIFY INDUSTRY;
        INDEX CREATE COMPANY=(INDUSID CUSIP);
        CONTENTS DATA=INDUSTRY;
RUN;
```

The CONTENTS command prints the contents information for the SAS data set COMPSTAT.INDUSTRY.

❷ **Indexing the newly created SAS data set.** The following example uses the option on the DATA command to create a composite index named COMPANY for a new SAS data set COMPSTAT.INDUSTRY. Two variables, INDUSID and CUSIP will be used as key variables.

```
DATA COMPSTAT.INDUSTRY(INDEX=(COMPANY=(INDUSID CUSIP)));
```

With the release of SAS 6.07, you can use the DATA command to create both a simple index and a composite index at the same time. For example, the following statement uses the variables SSN, CITY and STATE to create a simple index named SSN and a composite index named CITYST.

```
DATA EMPLOYEE(INDEX=(SSN CITYST=(CITY STATE)));
```

If a key variable has missing values, you can omit them from the index. The following statement create a simple index named SSN. This statement also uses the variables CITY and STATE to create a composite index named CITYST. The NOMISS option is used to exclude all observations with missing values.

```
DATA EMPLOYEE(INDEX=(SSN/NOMISS CITYST=(CITY STATE)/NOMISS));
```

Eliminating Disk Space Problems by Compressing Your SAS Data Set

SAS 6.07 is more efficient in conserving disk space for uncompressed SAS data sets than any previous releases. However, it is recommended that you use the COMPRESS option in your SAS program. By using this option in Release 6.06 , you could reduce your disk space usage down to 30% for your SAS data sets. Release 6.07 provides even faster and smaller disk compression and decompression operations. There are several ways to use this COMPRESS option in your SAS program; two of them are mentioned on the following page.

❶ **Using this option on the first line in your SAS program.** This can be accomplished by putting the following command on the first line of your SAS program:

OPTIONS COMPRESS=YES;

By placing this command on top of your SAS program, all temporary and newly created SAS data sets will be compressed.

❷ **Using this option with the DATA command.** This can be done by specifying the COMPRESS statement in parentheses after the SAS data set name in a DATA step statement, for example:

DATA NEW(COMPRESS=YES);

By using this approach, you will have only that newly created SAS data set compressed. Other SAS data sets will be saved in the default format which is uncompressed.

When you create a new, compressed data set, a note is written to the log indicating the percentage of reduction obtained from compressing the data set. The compression percentage is obtained by comparing the size of the compressed data set with the size of a noncompressed data set of the same page size and record count. ■

Researching the National Archives

By Theodore J. Hull, Center for Electronic Records (tif@cu.nih.gov)

This is an edited message that was posted to Public-Access Computer Systems Forum (LIBPACS@UHUPVM1)

The Center for Electronic Records of the National Archives and Records Administration has recently revised its *General Information Leaflet 37*, "Information About Electronic Records in the National Archives for Prospective Researchers." Included in GIL

Please see ARCHIVES on page 12.

CRSP Data Information

By Panu Sittiwong, Research and Statistical Support Manager (panu@unt.edu)

The following information applies to CRSP tapes with data ending December 1991. Since the Daily Return and OTC data are used frequently, they are now stored as cataloged data sets on an HDS-8083 mainframe system disk. The sample program is stored on a public disk in both MUSIC/SP and VM/CMS. All subroutines are stored on the ACAD system disk. In addition, there are also the precompiled modules for all the subroutines stored on an OS/MVS disk. The precompiled modules will help speed up your program if you haven't made any modifications to the subroutines. The sample program below shows how you can use the load module.

CRSP Daily Return File (Data are stored on Disk)

Contents	DSN	RECFM	LRECL	BLKSIZE
Sample Program	USER.A000.CRSP.SAMPROG	FB	80	4000
Sample CUSIP Program	USER.A000.BISTKC.SAMPROG	FB	80	4000
Sample Name List Program	USER.A000.NAMLST.SAMPROG	FB	80	4000
Includes Subroutines	USER.A000.CRSP.INCLUDE.PDS	FB	80	4000
Access Subroutines	USER.A000.CRSP.ACCESS	FB	80	4000
Utility Subroutines	USER.A000.CRSP.UTILITY	FB	80	4000
Subroutine Object Library	USER.A000.CRSP.OBJLIB	U	0	32760
Calendar File	USER.A000.DRET.CALENDA	VB	48	32760
Data File	USER.A000.DRET.DATA	VB	29704	32760

NASDAQ File (Data are stored on Disk)

Contents	DSN	RECFM	LRECL	BLKSIZE
Sample Program	USER.A000.CRSP.SAMPROG	FB	80	4000
Sample CUSIP Program	USER.A000.BISTKC.SAMPROG	FB	80	4000
Sample Name List Program	USER.A000.NAMLST.SAMPROG	FB	80	4000
Includes Subroutines	USER.A000.CRSP.INCLUDE.PDS	FB	80	4000
Access Subroutines	USER.A000.CRSP.ACCESS	FB	80	4000
Utility Subroutines	USER.A000.CRSP.UTILITY	FB	80	4000
Subroutine Object Library	USER.A000.CRSP.OBJLIB	U	0	32760
Calendar File	USER.A000.NASDAQ.CALENDA	VB	48	32760
Data File	USER.A000.NASDAQ.DATA	VB	19268	32760

CRSP Monthly Index and Return File (Data are Stored on Tape — TMS Vol. 300820)

(File #) Contents	DSN	RECFM	LRECL	BLKSIZE
(1) Tape Control Info.	USER.AC09.UMZCD1	FB	80	1600
(2) Copyright	COPYRITE.CRSP1992	FB	80	80
(3) Binary Sample Program	BISTK	FB	80	4000
(4) Character Sample Program	CHSTK	FB	80	4000
(5) CUSIP Input Sample Prog.	BISTKC	FB	80	4000
(6) Name list Sample Program	NAMLST	FB	80	4000
(7) Include Files	INCLUDE	FB	80	4000
(8) Access Subroutines	ACCESS	FB	80	6160
(9) Utility Subprograms	UTILITY	FB	80	6160
(10) Include Files PDS	INCLUDE.PDS	VS	6156	6160
(11) Access Subroutine PDS	ACCESS.PDS	VS	6156	6160
(12) Utility Subprograms PDS	UTILITY.PDS	VS	6156	6160
(13) Object Library PDS	OBJLIB.PDS	VS	32760	32760
(14) Calendar/Indices	MSIC#91C	VB	48	32760
(15) Data File	MZC#91C	VB	26836	32760

Please note also that since 1990, CRSP has changed the format of all tapes. Hence, you may need to get a new sample program source code and subroutine in order to process the new data (post-1990). The next section illustrates the retrieval of the CRSP sample program and all the subroutines.

Steps in Using CRSP Datasets

The following are steps that you need to follow in order to access CRSP data sets.

❶ Copy the sample program from the public library to your library. This can be done by issuing a command from the MUSIC/SP ***GO** prompt: **COPY CRSP.FORT** ***filename*** (where ***filename*** is any valid MUSIC/SP file name).

On CMS, the command is **COPY CRSP FORT *** ***fn ft*** **A** (where: ***fn*** and ***ft*** are any valid CMS filename and filetype).

❷ Use the editor to modify the sample program to your needs. At a minimum you will need to modify the JCL to correspond to your User-ID and the data set that you want to use. The input data are defined in the program with **//GO.FT10F001 DD** and **//GO.FT11F001 DD.**

❸ Determine whether or not you need to modify the subroutines. If you need to make modifications to the subroutines, you will need to retrieve them to your library using the **IEBGENER** program. The sample program is included in Figure 2.

❹ Submit the program for execution.

Sample Program Setup

CRSP requires that you use a FORTRAN program to access the data. Thus, the sample program and the subroutines are required in order to access the data. The program, however, is for general purpose use; you must add your own codes in order to obtain particular data items. The following sample shows a job stream for CRSP data processing. Please call Panu Sittiwong in Academic Computing Services if you need more assistance or information. He can be reached at (817) 565-2140, Monday through Friday 8:00 a.m. to 5:00 p.m.

Figure 1: Sample Program to Access CRSP Data Sets

```
/INC OSJE
RETURN
//IDNNCRSP JOB (idnn,8,20),'Your Name',CLASS=J,PASSWORD=mvspw
// EXEC FORVCLG
//FORT.SYSLIB DD DSN=USER.AC09.CRSP89.INCLUDE,DISP=SHR
//SYSIN DD *
//*************************************************************
//*  FORTRAN PROGRAM
//*************************************************************
//LKED.SYSLIB  DD DSN=USER.A000.CRSP.OGJLIB,DISP=SHR
//              DD DSN=SYS1.VFORTLIB,DISP=SHR
//GO.SYSIN      DD *
//GO.FT10F001   DD DSN=XXXXXXX,DISP=SHR
//GO.FT11F001   DD DSN=XXXXXXX,DISP=SHR
```

Where: FT10F001 refers to the Calendar file. The DSN is USER.A000.DRET.CALENDA for the Daily Returns data, and it is USER.A000.NASDAQ.CALENDA for the NASDAQ file.

FT11F001 refers to the Data file. The DSN is USER.A000.DRET.DATA for the NASDAQ file.

Since the Monthly Data are stored on tape, the //GO.FT10F001 DD and //FT11F001 DD will be as follows:

```
//GO.FT10F001  DD DSN=MSIC#91C,UNIT=TAPECR,
//                VOL=SER=300820,LABEL=(14,SL),
//                DISP=(OLD,PASS)
//GO.FT11F001  DD DSN=MZC#91C,UNIT=AFF=FT10F001,
//                VOL=SER=300820,LABEL=(15,SL),
//                DISP=(OLD,KEEP)
```

Figure 2: Sample IEBGENER Program

```
/INCLUDE OSJE
RETURN
//idnnENER JOB (idnn,:05,1,9999),'Your Name',PASSWORD=mvspw
// EXEC PGM=IEBGENER
//SYSPRINT DD SYSOUT=A
//SYSIN DD DUMMY
//SYSUT2 DD SYSOUT=B
//SYSUT1 DD DSN=xxxxxxxxx,DISP=SHR
```

Where: xxxxxxxxxx is a data set name.■

ARCHIVES continued from page 10.

37 are brief descriptions of the National Archives and the Center for Electronic Records, and how we obtain and maintain the records in our custody (Attitudinal Data, Demographic Data, Economic and Financial Statistics, Education Data, Health and Social Services Data, International Data, Military Data, and Scientific and Technological Data).

If anyone is interested in receiving a copy of *GIL 37*, please contact Theodore J. Hull at the Center for Electronic Records, National Archives (Internet: tif@cu.nih.gov or BITNET: tif@nihcu) Phone: 202-501-5579 or write to the Center for Electronic Records (NNX), ATTN: Reference Services, National Archives, Washington, DC 20408. ■

Formatting Theses and Dissertations in WordPerfect

By James Yarbrough, Research and Statistical Consultant (AC35@UNTVM1)

WPThesis is a set of WordPerfect macros that are designed to assist graduate students with the formatting aspects of writing theses and dissertations. WPThesis macros perform a wide variety of functions from numbering pages, setting margins, creating a table of contents, creating chapter headings, lists of tables, references, etc., all in the official format set forth by the graduate school. The macros can be incorporated into Version 5.1 of WordPerfect.

WPThesis was originally developed at Auburn University. ACS has acquired it from them and we are currently revising it so that the formatting operations are in accord with UNT guidelines for writing theses and dissertations. The revised version of WPThesis should be available to all UNT graduate students early this fall. Graduate students can contact ACS early in the fall semester for more information on how to obtain a free copy of WPThesis. ■

The Network Connection

By Billy Barron, VAX/UNIX System Manager (billy@unt.edu)

This column is a continuing feature of ***Benchmarks*** *intended to present news and information on various aspects of wide area networks.*

Using Wide Area Networks for Research

by Billy Barron, VAX/UNIX System Manager (billy@unt.edu)

Wide Area Networks (WANs) are very powerful tools that can be used by researchers. Knowing where to look on WANs for information can save researchers hours of library research or allow unique data to be found that may not otherwise be available.

On-line Libraries

Many university library catalogs are available over the Internet. Currently, they number in excess of 350 sites and provide a variety of services for researchers on the UNT campus.

First, they give researchers the ability to search other libraries for books that they can borrow through InterLibrary Loan (ILL). Also, this access allows them to check the status of a book before submitting an ILL request. Secondly, researchers have the ability to check the holdings of another library before going to research there. For example, a researcher could check the UTD library to see if UTD has the needed materials before making the trip down there. The on-line library systems also allow researchers to work at almost any time of the day instead of just the times when the library is open. Researchers can additionally check the holdings of libraries at a colleague's university library while working on collaborative projects. Finally, the online libraries allow for the easy building of comprehensive bibliographies.

A related service is CARL (Colorado Association of Research Libraries) UNCOVER2. UNCOVER2 is a database of journal and magazine articles. Searching the database is free to researchers, but the retrieval of the text of the articles does cost money. CARL UNCOVER2 is currently available at the reference desk in the Willis Library.

A handout listing the libraries in the North Texas region is available in the Computing Center, ISB 119, free of charge. The regional libraries can be accessed via a menu on the public PCs at the reference desk in the Willis Library. A more detailed document "UNT's Accessing On-line Bibliographic Databases" is available for reference in most of the General Access Labs. The document is also available via anonymous FTP on ftp.unt.edu in the pub/library directory with the file names **libraries.***.

The information is also available in hypertext format. The MS-DOS hypertext program is called HYTELNET. It is available for anonymous FTP on ftp.unt.edu in the pub/library/hytelnet/pc directory. Better yet, HYTELNET can be accessed by just typing **hytelnet** at the VAX or Solbourne prompt. It is eventually our hope to move the HYTELNET information into the NT Gopher system (see the article "NT

Gopher: Academic Information Service" in the June 1992 issue of *Benchmarks*, page 8), so there will be one interface for all of these services.

Communication with Colleagues

Instead of playing telephone tag with expensive long distance phone bills or US mail with its long delays, you can send electronic mail to colleagues around the world. Most of the time electronic mail is delivered in a matter of minutes. For more information about electronic mail, see the June 1992 issue of *Benchmarks*.

USENET News/Mailing Lists

Most disciplines have USENET newsgroup(s) and/or BITNET/Internet mailing lists associated with them. The newsgroups and mailing lists allow researchers to keep abreast of current events in their field. Also, researchers can ask small research related questions to the list and receive back responses. However, it is very important to verify information gathered this way. Many of the people on the lists are not experts in the field and tend to just say the first thing off the top of their head without verifying its accuracy. On the other hand, most lists have at least a few highly knowledgeable, highly qualified people. (See "The Network Connection" in the March/April 1992 *Benchmarks* for more information).

Electronic Journals

A few journals today are being published in only electronic format. The distinction between an electronic journal and a mailing list is that a journal is similiar to a magazine and a mailing list is a discussion oriented forum. We are placing many of the free electronic journals into NT Gopher. You can find them under the "Newsletters & Journals" section in NT Gopher. If you are aware of any electronic journals (not mailing lists) that you would like us to add to NT Gopher, send mail to gopher @unt.edu. Related to this, many universities publish their research reports in electronic format.

List of the Month

Each month we will highlight one BITNET, Internet, or USENET Special Interest Group (SIG) mailing list. This month's list...

This month's list is actually a USENET newsgroup. It can be accessed via ANU on the VAX or nn on the Solbourne. This information is also available in NT Gopher in the "Online Databases & Docs/Science/NSF" section.

info.nsf.grants

All sorts of useful information can be found in this USENET newsgroup. For example, announcements of new posts to the NSF bulletin board, STIS, are posted on a regular basis. The *NSF Bulletin* is also posted there as it is published. Other types of items such as "Summary of the House Appropriations Markup FY 1993 NSF Budget Request" can also be found in this newsgroup.

To access ANU NEWS on the VAX type **ANU** from the $ prompt. To access nn on the Solbourne, type **nn** at the % prompt. Help screens exist for each of these conferencing systems on their respective host systems. Additionally, documentation on ANU News is available in the Computing Center, ISB 119. Ask for "ANU News User's Guide" and "ANU News Reference Guide." More information about NT Gopher can be found in the article "NT Gopher: Academic Information Service" in the June 1992 issue of *Benchmarks*, page 8. ■

Various databases

All over the Internet, people have built databases on various subjects. They use a variety of software to run the databases including WAIS, Gopher, and WWW. Here at UNT, we are attempting to use Gopher as a common front end to all these resources. Databases are constantly being added to our Gopher system under the "Online Databases & Docs" section.

One of the disciplines that has an excellent set of databases is Biology. Biology databases that are available through NT Gopher include Genbank, PIR, and Drosophila. Some of the more interesting non-Biology databases available include Long-term eco-

logical research, NSF grant information, Supreme Court decisions, and song lyrics. There are many more than are listed here and more are being added on a regular basis.

Conclusion

The WANs, especially the Internet, are valuable research tools. If you have any questions about what is available or how to access it, feel free to contact Academic Computing Services (ISB 119, 565-2324) for more information. We do not have a complete list of resources available (nobody in the world does), but we know how to look for a particular resource and tell you if it is currently available or not. Also, we will attempt to keep all of the known resources available through NT Gopher.■

FEDIX:

An On-Line Information System for Colleges, Universities, and Other Research Organizations

This announcement was posted to lib_hytelnet@sask.usask.ca.

FEDIX is an on-line information service that links the higher education community and the federal government to facilitate research, education, and services. The system provides accurate and timely federal agency information to colleges, universities, and other research organizations.

FEDIX provides daily information updates on:

- Federal **education** and **research programs** (including descriptions, eligibility, funding, deadlines).
- **Scholarships, fellowships,** and **grants**
- Available used government **research equipment.**
- New funding for specific research and education activities from the *Commerce Business Daily*, *Federal Register*, and other sources.
- **Minority assistance** research and education programs.
- **News & current events** within participating agencies.
- **General information** such as agency history, budget, organizational structure, mission statement, etc.

Participating Agencies

Currently FEDIX provides information on 10 federal agencies. The Department of Energy (DOE), Office of Naval Research (ONR), National Aeronautics and Space Administration (NASA), Air Force Office of Scientific Research (AFOSR), and Federal Aviation Administration (FAA) are providing comprehensive education- and research-related agency information, while the National Science Foundation (NSF), Department of Housing and Urban Development (HUD), Department of Commerce (DOC), Department of Education (DOEd) and National Security Agency (NSA) are providing minority assistance information, exclusively. Additional government agencies are expected to join FEDIX in the future.

MOLIS, the Minority On-Line Information Service, <OTH059> complements FEDIX by providing current information about Black and Hispanic colleges and universities. Data Line: 301-258-5015 or 800-626-6547. Office: 301-975-0103.

Accessing FEDIX

To access FEDIX from a host system at UNT type **telnet fedix.fie.com** or **telnet 192.111.228.1** Log in as **fedix.** A help line is available at 301-975-0103.■

Using the SPSS Macro Facility

By James Yarbrough, Research and Statistical Consultant (AC35@UNTVM1)

The SPSS macro facility is a powerful set of operations that offer the statistical programmer options for dealing more effectively with a variety of programming situations. Perhaps the most significant feature of SPSS macros are their ability to allow the programmer to execute a series of similar command lines in a loop fashion, rather than having to include a number of redundant blocks of coding. This article will concentrate on this feature of the SPSS macro facility. Besides the looping function, SPSS macros can also be used to group sets of variables, or to produce output from several SPSS procedures with a single command statement.

All macros consist of two parts: the *macro definition* and the *macro call*. In the macro definition, the macro is given a name (the *macro name*), arguments are defined and named, and the body of the macro is included. Macro definition begins with the DEFINE command followed by the macro name. Arguments then follow (if needed), and their type is established. The body of the macro is next, consisting of SPSS commands—typically, arguments will be referred to in the macro body, although this is not required. Finally, the end of the macro is marked with the !ENDDEFINE command.

Pursuant to macro definition, the macro is "called." A *macro call* is achieved simply by including the macro name (as if it were a command). Also as part of the macro call, arguments are given specific values (if there are any arguments). The values given the arguments are substituted in each place that the argument is found in the body of the macro. Then the body

of the macro is executed using these values. Subsequent macro calls (or *macro runs*) will run the block of statements (macro body) with different values for the arguments. In this way, redundant coding can be reduced—a block of commands can be run over and over, changing slightly on each occasion, simply by writing the code once (including macro arguments) and then including multiple macro calls in which arguments change values for each run.

The following is an example of the type of macro described above:

```
define cycle (fileout = !tokens(1)
              / n = !tokens(1))
get file = 'alldata sysfile a'
sample !n
save outfile = !fileout
!enddefine
cycle   n = .10
        fileout = 'small file a'
cycle   n = .33
        fileout = 'medium file a'
cycle   n = .75
        fileout = 'large file a'
```

The above macro is named "cycle." It has two arguments, "fileout" and "n," both of which are one string "tokens" (more on this later). The body of the macro consists of the statements from "get file" to "save outfile" (inclusive). Note that when arguments are referred to in the body of the macro, they are prefixed with an exclamation point (!). This program will execute the body of the macro three times, as the macro is "called" three times. The first time through, n is .10 and fileout is 'small file a'. In this first run, a random sample of 10% of the cases is drawn and saved as a system file called 'small file a'. In the second run (call), a random sample of approximately one third of the total number of cases is drawn and saved as 'medium file a. In the final run, a random sample of one half of the cases is drawn and saved as 'large file a'.

Of course, there are a great number of variations possible on the above theme. In all cases, the macro is defined and then called one or more times. The following is another simple example:

```
get file = 'data sysfile a'
define group (set = !charend ('/'))
frequencies variables = !set
descriptives variables = !set
!enddefine
group set = age sex educ /
group set = opinion taste /
```

Here, the system file is opened, and the macro defined. The macro in this example is named "group." The body of this macro consists of the *frequencies* and *descriptives* statements. Two macro calls are specified. In the first macro call, set has three values: age, sex, and educ. In the second call, set only has two values, opinion and taste. In each macro call, the argument is *expanded* in the macro body. Hence, "!set" becomes "age sex educ" in the first run (call), and becomes "opinion taste" in the second run. Therefore, frequencies and descriptives will first be produced for age, sex, and educ, and then, in the second run, they will be produced for opinion and taste.

The ability to have a variable number of values for the argument (set) is due to the use of !charend as the type of argument. That is, !charend ('/') specifies that all values included after "set =" in the macro call, and up to the "/" will be included as values of the argument, set. When "token(n)" is chosen as the argument type, the number of values for the argument is fixed and is equal to n.

Two other types of arguments are available, both of them allowing a variable number of values for the argument. One type is !enclose ('startsym', 'endsym'). Startsym stands for "start symbol." Values for the argument are chosen beginning with the first values to follow the start symbol, whatever that symbol is chosen to be. Then, values are chosen until the "end symbol" (endsym) is reached. The final type of argument is written as !cmdend. !cmdend instructs SPSS to assign to the argument all of the text remaining in the macro call up to the next command. Hence, !cmdend can only be used for the last argument in the macro call—if it is not, subsequent arguments will be read as text in the previous argument.

It is not required that macros have arguments. At first this may seem strange, but consider the following example:

```
define update ()
split file by group
frequencies variables = reactime accuracy
  /histogram
descriptives reactime accuracy
list
split file off
regression variables = group reactime accuracy
  /dep = accuracy
  /enter
!enddefine
update
compute reactime = sqrt(reactime)
compute accuracy = sqrt(accuracy)
update
compute reactime = lg10(reactime * reactime)
compute accuracy = lg10(accuracy * accuracy)
update
```

In this example, a macro (with no arguments) is defined and then executed. Frequencies, descriptives, a listing, and a regression analysis are conducted in this first run (call) of the macro (called "update"). Next, the values of reactime and accuracy are transformed, and the macro is called again. During this call, the same procedures as before are carried out on the transformed variables. Finally, the process occurs once more—the variables are transformed and the tests are reconducted.

Hopefully, this article will give the reader some idea of the potential usefulness of the SPSS macro facility. By no means have all of the possibilities been explored. The macro facility is a very flexible tool, and allows for much creativity on the part of the individual programmer. In using the macro facility, SPSS programs may gain a certain subtlety and sophistication—redundant coding is reduced to a minimum, and in the process the programmer is freed to concentrate on more pressing programming issues. ■

Violations of Assumptions:

A Comparison of Parametric and Nonparametric Alternatives

By James Yarbrough, Research and Statistical Consultant (AC35@UNTVM1)

A typical student receiving a bachelor's degree in the social sciences will have taken a single course in statistics. Approximately 90% of this course will be devoted to parametric statistical techniques—often the percentage is higher. Nonparametric statistical techniques are either briefly studied, mentioned in passing, or (most often) never mentioned at all. While graduate students may take more research and statistics courses, the total percentage of time dedicated to parametric versus nonparametric approaches is similar to that for undergraduate students. In fact, upon completion of their studies, the vast majority of both undergraduate and graduate students are unable even to make the distinction between the two approaches. Taken together, the procedures most students (and graduates) consider as their statistical repertoire usually consist solely of parametric techniques (i.e., the t-test, anova, multiple regression), with the occasional exception of the chi square test. These facts, together with the acknowledgement that assumptions of parametric procedures are not fully met in most social science research situations, point to the need for a greater awareness of nonparametric techniques and of the effects of violations of assumptions of parametric tests.

Parametric -vs- Nonparametric?

What is the distinction between parametric and nonparametric statistical tests? Generally, parametric statistical tests are those that require at least interval level measurement and a sampling distribution that is normal in shape. Some parametric tests incorporate more assumptions. Parametric tests typically involve the calculation and manipulation of "parameters," such as the mean and standard deviation, which lack meaning if the underlying variables are not measured at at least the interval level. Hence, interval level measurement is required if one is to place faith in the validity of the parameters used in parametric analysis. Further, parametric inferential procedures generally make probabilistic statements based on the assumption that the underlying distribution (or at least the sampling distribution) is normal. In the social sciences, neither of these assumptions are often met. Most parametric procedures require that even more assumptions be met, such as homogeneity of variance between groups (t-test, anova).

Like parametric approaches, nonparametric approaches vary in the number and kind of assumptions that must be made. However, collectively, they never require more than ordinal level measurement and they rarely make assumptions about the shape of the underlying distribution. These qualities should appeal to the social scientist, for rarely is more than ordinal level measurement acquired and few frequency distributions are truly normal. While most parametric tests require only that the sampling distribution be normal, the small sample sizes typical of most social research deny the viability of meeting even this assumption.

There are many reasons for the bias toward the teaching of parametric approaches. First and foremost, when all assumptions are met, parametric approaches are the most "powerful" alternative. The power of a statistical test is inversely related to the sample size required to obtain a given significance level. This power superiority is to be expected, for if the more stringent assumptions associated with parametric statistics are met, more is necessarily known about the data. The mean, or average, for example, is unique as a measure of central tendency in that it uses value information from all observations in its formulation. Knowledge of a normal distribution also enables the researcher to make stronger conclusions about differences and associations. However, if the assumptions are not met, parametric procedures may produce very misleading results—in such cases, nonparametric approaches are often more valid and/or more powerful.

Second, there is a general consensus in the research community that "parametric statistical techniques are *robust* to mild violations of assumptions." However, this general acknowledgement is perhaps too often relied upon. As Charlotte Royeen (1986) states in a paper to the U.S. Department of Education, "the claim of robustness to violations of assumptions appears to have been overgeneralized beyond highly specified situations (usually delineated from simulation studies) to almost all real life situations and data sets. Thus almost all violations of assumptions,

often ill defined, are tolerated in the name of robustness." This article will attempt to summarize more precisely what some research has found about the impact of various levels and types of violations of parametric assumptions.

Parametric approaches are the most powerful as long as assumptions are met. This is the primary reason for their popularity. Additionally, to a large degree, the researcher has control over whether or not the assumptions are met. Large samples can generally guarantee a normal sampling distribution. And interval level measurement, while often not theoretically possible for many of the variables of interest to the social scientist, can be "approached" to the degree that parametric analysis regains some meaning and is thus more valid. For example, while grades or IQs can never be more than ordinal level variables, sophisticated measuring strategies can allow these measures to "approach interval level." In order to do so, measuring instruments "operationalize" the variable in a comprehensive and discriminating way. Hence, we feel justified in calculating final scores as means of test scores, and then using these scores as dependent variables in regression analyses, etc. Measurement of an ordinal level variable that approaches interval level is often referred to as "ordered-metric" measurement. Such measurement allows not only an acknowledgement of the order of scores, but also some notion of the relative distances between scores.

Some studies have been conducted to assess the effects of various violations of parametric assumptions and the relative performance of parametric versus nonparametric tests under those conditions. Stephen Olejnik and James Algina (1984) did a comparative analysis of parametric covariance with five nonparametric alternatives, examining the effects of violations of parametric assumptions. As expected, when all parametric assumptions were met, parametric ancova was the more powerful procedure. However, they state that under such conditions, "very little power is lost when a nonparametric procedure is used." One of the nonparametric alternatives to the analysis of covariance that was considered was that of Quade. The asymptotic relative efficiency of the Quade statistic was found to be between 87 and 95%.

In another comparative study, cited by Olejnik & Algina, Conover and Iman looked at the effect of non-normal distributions on the efficiency of parametric and nonparametric analysis of covariance. When the distribution was normal, the parametric approach held a 7% power advantage. However, large power differentials, favoring the nonparametric approaches, were found when the distribution was non-normal. When the distribution was lognormal, a 57% difference in power favoring the nonparametric approaches was found. When the distribution was exponential, the power difference was observed to be 30%. When the distribution was Cauchy, the power advantage was 25%. Abunnaja (cited by Olejnik & Algina) found differences in the same directions, but considerably less extreme. He also tested the effect of a uniform distribution, and found the parametric approach held a 3 to 5% power advantage.

These results indicate that nonparametric approaches give up little when the distribution is normal, but can be considerably more powerful when the distribution departs from normality. Further, when the decision between the two, in terms of power, is a toss up, one should realize that nonparametric approaches can allow for a more valid generalization to non-normal populations. It is usually possible to inspect the nature of the conditional distribution by examining the sample frequency distribution Y_{ij} or ε.

Simulation studies such as those above tend to disregard violations of assumptions concerning appropriate levels of measurement. This is because the effects of such violations are not easily measured or tested. That is, rather than causing a change in the power of a test, violations of measurement level assumptions tend to detract from the "meaning" of the measurement, and hence from the validity of the result. Suffice it to say that parametric tests conducted on less than interval level data are suspect, and that any time this occurs the researcher is obliged to state the violation in reports along with findings. As mentioned above, ordinal level data can "approach" interval level through the sophisticated use of a carefully designed measuring instrument. The measuring instrument should be designed so as to measure the variable of interest in a comprehensive and discriminating fashion. The resultant measure should describe not only the order of scores but also give some approximation of the relative differences between scores. Such measures are often referred to as "ordered-metric."

Conclusion

While parametric procedures are the most well known, the researcher does have at his/her disposal a wide array of nonparametric procedures. There is generally a corresponding nonparametric procedure for all applications in which one might consider using a parametric test. Computer statistical packages usually incorporate a fair selection of these procedures. In SPSS, most of the nonparametric procedures are included (as subcommands) in the NPAR TESTS command. For example, the nonparametric Mann-Whitney test is the complement of the parametric t-test. Nonparametric correlation procedures are included in the NONPAR CORR command in SPSS. Also, the chi-square statistic may be requested in the CROSSTABS procedure in SPSS.

References

Olejnik and Algina. "A Review of Nonparametric Alternatives to Analysis of Covariance." University of Florida, 1984.

Royeen, Charlotte B. "A Comparison of Parametric Versus Nonparametric Statistics," U.S. Department of Education, Washington, D.C., 1986. ■

Assistant to Associate V.P. for Computing Hired

Susan Pierce, Computer Systems Manager for the School of Community Service, has been hired to replace Stanley Sawyer as the Assistant to the Associate Vice President of Computing. Sawyer, who retired this past year, has been working in the position part-time. Pierce and Sawyer will continue to share the position until January 1993, when Pierce will assume the duties full-time. She will continue in her position, part-time, as Computer Systems Manager for the School of Community Service until that time. ■

Computing Center Staff Activities

Academic Computing Services

An article "The Left Hand Doesn't Know What the Right Hand is Doing" by **Billy Barron**, VAX/UNIX Systems Manager, appeared in the "Martian Packets" column of *Matrix News* (April 1992, Vol. 2 No. 4, p. 5). In the article, Barron discusses the need for greater cooperation in providing information resources on the Internet. ■

Presidential E-mail

Want to chat with your favorite candidate? You might get to if you send mail to the following addresses:

- ❑ **Jerry Brown:** 75300.3105@compuserve.com **or**

Please see CANDIDATE on page 19

Information Resources Council News

Minutes provided by Sue Harrison, Recording Secretary

Information Resources Council Members: *Philip Baczewski, Computing Center (ex-officio); Dave Barker, TCOM - Physiology; Bill Buntain, Computing Center (ex-officio); Cengiz Capan, College of Business; Carolyn Cunningham, Financial Aid; Jim Curry, Micro Maintenance (ex-officio); Stephen Farish, College of Music; Jean Schaake, College of Arts & Sciences; Chuck Fuller, Business Services; Don Grose, UNT Libraries; Richard Harris, Computing Center (ex-officio); Steve Miller, Personnel; Tom Newell, Telecommunications; Don Palermo, Admissions; Sue Pierce, School of Community Services; Paul Schlieve, Computer Education & Cognitive Systems; Neal Tate, Political Science; Ray Vondran, Library and Information Sciences (Chair); Sue Harrison, Computing Center (Recording Secretary).*

Tuesday, June 16, 1992

The minutes of the May 7, 1992 meeting were approved, with the addition of one guest's name, David Hartman, to the list of those present.

Neither Bill Buntain nor Paul Schlieve were available to report on Network Directions; therefore, news on this issue will be deferred to the July meeting.

Philip Baczewski highlighted some of the discussion from the Unix Advisory Committee meeting of May 13, 1992. It was the original intention of this Advisory Committee to make the Solbourne computer available for providing a high performance computing environment; and the Committee believes it should remain so. However, because of its capabilities, they have decided that instructional use could be made available in a limited capacity. The Advisory Committee proposes that the following statement be added to the Academic Computing Services Policies and Procedures document:

> The Solbourne system's primary purpose is to provide a high-performance computing environment for research computing. Faculty, staff, and graduate students may be granted a User-ID on this system. Undergraduate User-IDs and Classroom User-ID sets may also be allocated for the Solbourne, however, the VAX/UNIX Systems Manager will periodically review class and undergraduate usage in order to maintain system performance consistent with the purpose stated above.

A vote will be taken on this recommendation at the July meeting.

Philip Baczewski distributed a proposed FY93 HEAF Computing Budget for IRC members to review and discuss at the July meeting.

Ray Vondran reported that the issue of additional space for the Microcomputer Maintenance Shop was again brought before the IRC Steering Committee at its most recent meeting. Dr. Brownell stated at that meeting that he would talk with Jack Davis about this issue. When asked by Dr. Vondran if Jim Curry had heard anything about additional space for his shop, he reported that he had not.

Chairman Vondran also noted that a report from the Instructional Technology Task Force is being forwarded to the President and to Deans for their comments and final approval. Until that final approval is given, the IRC will defer appointing an Instructional Technology Standing Committee.

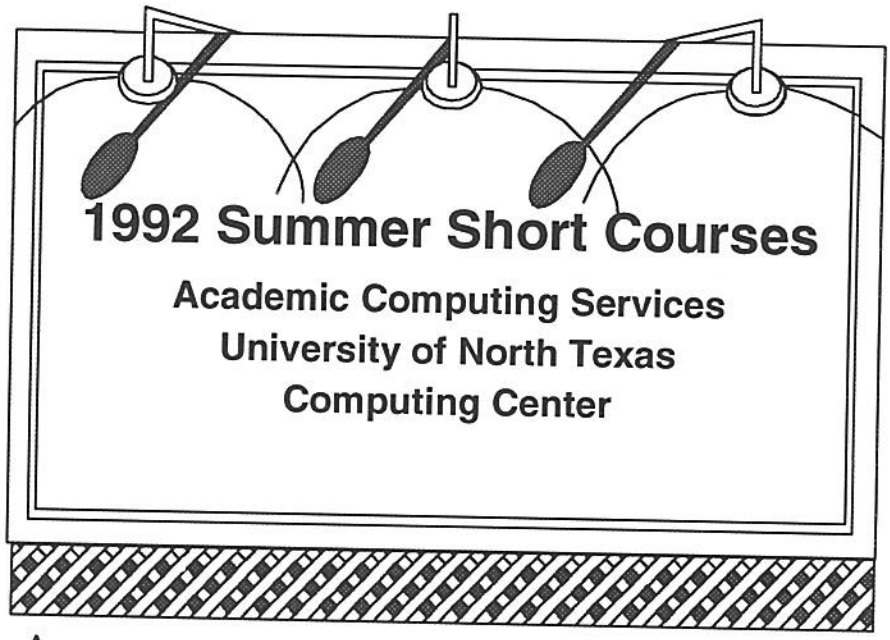
Philip Baczewski reported that the General Access Lab Committee met on June 2, at which time their main focus was a demonstration of a check-in program that is being used in all of the General Access Labs across campus now. This program was coded and implemented by Eric Lipscomb of the Computing Center, and will work with data from the SIMS and HRMIS systems. The committee is working with departments now to get proper permission to use the data on these two systems. There are some freedom of information issues that need to be addressed and a policy to be written. A subcommittee has been appointed to look at these issues.

Another proposal made at the GALC meeting was the possibility of utilizing the back part of the General Access Lab in Chilton Hall, room 255 for faculty/staff and student training during nonpeak use times of that facility. A second subcommittee was appointed to work with Steve Miller in determining what would be needed to equip room 255 for use as a faculty/staff/student training facility as well as explore issues of the appropriateness of the proposal.■

CANDIDATE continued from page 18

brown92@igc.org or brown 92@well.sf.ca.us

- ❑ **George Bush:**
- ❑ **Bill Clinton:** 75300.3115@compuserve.com
- ❑ **Andre Marrou:** 75300.3114@compuserve.com
- ❑ **Ross Perot:** No address at this time.■



Academic Computing Services is offering the following short courses for the 1992 summer sessions. Please preregister to attend (a registration form can be found at the end of this issue). **A maximum of 10 people will be admitted to each of the courses held in ISB 110. A maximum of 8 people will be admitted to each of the courses held in ISB 123. Academic Computing Services reserves the right to cancel ANY course that has 5 people or fewer registered 3 days before the date of the course.**

PLEASE NOTE: Faculty and students have first priority to register for these classes. All people registering for hands-on (ISB 110) HDS, VAX and/or UNIX courses should have current User-IDs. Applications for User-IDs are available in the Computing Center main office (ISB 119).

STATISTICAL PACKAGE COURSES

1. **Introduction to SAS on UNIX** — A one-hour session to be held in the Science Library (ACS General Access Lab, ISB 110):
   - Monday, July 20: 4-5 p.m. Instructor: Panu Sittiwong

WIDE AREA NETWORK COURSES

1. **Introduction to the Internet — Prior knowledge of at least one of the following interactive operating systems is required: VAX/VMS, CMS, UNIX, MS-DOS, MAC.** A one and a half-hour session, to be held in the Academic Computing Conference Room (ISB 123):
   - Tuesday, July 21: 3-5 p.m. Instructor: Billy Barron
2. **Introduction to USENET — Prior knowledge of at least one of the following interactive operating systems is required: VAX/VMS, UNIX, MS-DOS, MAC.** A one-hour session to be held in the Computing Center Conference Room (ISB 123):
   - Thursday, July 23: 4:00-5:00 p.m.Instructor: Billy Barron ■

Calls For Papers, Conferences, And Other Items of Interest

We have received the following "calls" and announcements from various organizations.

Call for Papers, Proposals

- **6th Annual International Computer Virus & Security Conference, March 10-12, 1993, New York, NY Ramada and Marriott Marquis** — Papers on such topics as prevention, detection, and recover from viruses and other unauthorized usage; particulars of LAN, UNIX, crytography, and military use; computer crime, law, data liability, related contexts; US/international sharing of research & techniques; case studies of mainframe, PC &/or network security. December 18, 1992 is the deadline for submissions. Contact Judy Brand, Conference Chair, Box 6313 FDR Station, New York, NY 10150 for more information.

Workshops, Conferences, and Seminars

- **EDUCOM's Educational Uses of Information Technology (EUIT) Seventh Annual Snowmass Working Session, August 5-7, 1992, Snowmass, CO** — EUIT provides you with an opportunity to exchange ideas, get involved in EUIT's active project working groups, and the development of new ones. The theme this year is "Transforming Information Technologies into Educational Realities: You and I and EUIT." Contact EUIT@EDUCOM for more information.
- **EDUCOM'92, October 28-31, Baltimore, MD** — The theme of this year's conference is "Charting Our Course." More than 50 sessions focusing on today's most important issues for information technology professionals in higher education will be offered. An Educational Uses of Information Technology (EUIT) preconference working session will be held on October 27, 1992. For more information, send an interactive message or mail to LISTSERV@BITNIC containing **GET EDUCOM92 SEMINARS** or call 202-331-5375. For information about EUIT call 202-872-4200 or send an interactive message or mail to EUIT@EDUCOM.
- **The Network Services Conference 1992 (NSC'92), Pisa, Italy, November 3-5, 1992** — NSC '92 will focus on issues in providing services to customers, with special attention paid to the recent developments in new global high-level tools such as World-Wide Web, Prospero, Archie, Alex, Gopher, and WAIS. This conference will be of interest to network service providers and sophisticated users. Further information can be obtained by sending mail to NSCINFO@FRORS12.
- **CAUSE92, December 1-4, 1992, Loews Anatole Hotel, Dallas, TX** — The theme of this year's conference is "Grasping the Momentum of the Information Age." Besides the conference itself, ten preconference seminars are being offered on such topics as "The Virtual Library" and "Ethics in Cyberspace." Contact David J. Ernst (DERNST@UCSFVM) Phone: 415-476-6189 for more information.
- **A Consortium for Network Publication of Refereed Research Journals** — The University of Manitoba has received funding commitments to organize and hold an international conference to promote the establishment of a consortium of universities and learned societies to sponsor computer network publication of refereed journals. The consortium would be a nonprofit publishing cooperative intended to make use of the Internet as an important medium for the publication of scholarly research in any discipline. The conference is tenatively slated for the **Autumn of 1993** and will be held at the University of Manitoba. For more information, and to express your interests in the conference and consortium, contact the convenor of the University of Manitoba ad hoc Committee on Electronic Journals, Professor Larry W. Hurtado, Institute for the Humanities, 108 Isbister Bldg., University of Manitoba, Winnipeg, Manitoba, R3T WnW. Phone: 204-474-9114. FAX: 204-275-5781. E-mail: hurtado@ccu.umanitoba.ca ■

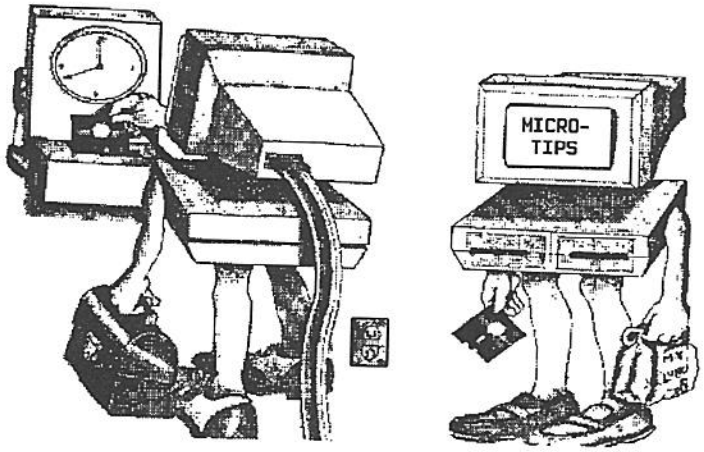
Accessing the UNT Library Catalog

By Billy Barron, VAX/UNIX System Manager (billy@unt.edu)

Over the past few weeks, access to the library catalog by users who were not using the Sytek LAN has been unstable. The problems included hardware, software and a lack of ports. On the evening of June 24, the library installed a network controller. During some tests, the new system was much more stable. Also, it supports 40 instead of 8 ports.

There are, however, a couple of minor changes on how to access the library. (This does not apply to Sytek LAN users.) When you connect to the library, you will get a prompt saying: **Type 'HELLO USER.LIB' to logon to the system:**. You must type HELLO USER.LIB at this stage to connect to the library. When you are finished you should type /QUIT to exit.

Users of the metro lines, 3300 lines, and telnet users should find this new system to be a drastic improvement over the previous system. ■



This column is intended to serve as a forum for sharing useful tips on making more productive use of microcomputers. If you have a tip that you feel may be of use to campus users, submit it to the ***Benchmarks*** *Editor for possible inclusion in a future issue.*

Supervising Your PC's Memory Managers

By Rich Caserta (caserta@jhuvm), The Johns Hopkins University

This article originally appeared in ***Checkbits*** *(Vol. 7, No. 9, pp.1-6), the newsletter of Homewood Academic Computing at The Johns Hopkins University. It has been slightly edited for publication in Benchmarks.*

In this article, I will discuss memory management for IBM compatible PCs. As PC software applications get larger and hungrier for memory in which to operate, it becomes critical that PC memory is optimally configured. A good memory configuration allows you to use newer applications and take advantage of their power. Even if your PC has megabytes of RAM (random access, or working, memory), it might not be using this RAM to your advantage; worse, it might not be using the extra RAM at all.

From a PC history lesson we will progress to a study of the current state of memory management. There, I will discuss methods you can use now to optimize your PC's memory use so you can work better. Then I will mention some coming developments that may relieve PC users of the burden of memory management forever.

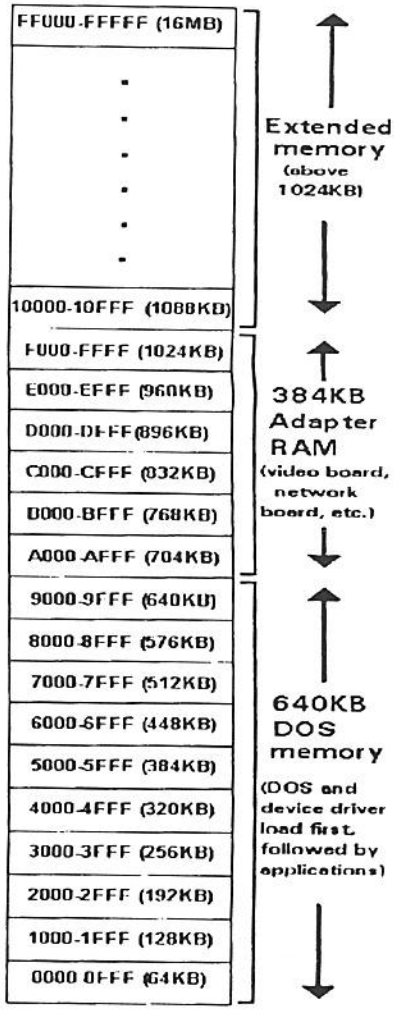
PC Memory Lane

When IBM introduced the first DOS PC in August 1981, PC memory management was no problem. The first PC's Intel 8088 microprocessor (CPU) could address 1024 kilobytes (KB), or 1 megabyte (MB) of random access, or "working," memory (RAM). In developing the PC and choosing DOS as its operating system, IBM engineers assigned the upper 384 KB of the 8088's 1 MB memory address space to hardware adapters (add-in boards), and the lower 640 KB to DOS and software applications. They believed these assignments would support any foreseeable evolution of application software or hardware adapters (Figure 1).

They had good reason to be confident. The first IBM Pcs were shipped with only 64 KB of RAM, one-tenth of the memory capacity designed for DOS and its applications. Most software programs available at the time did not fill the 64 KB workspace. The only hardware adapter that used RAM address space was the monochrome display adapter (MDA), occupying only 16 KB of the reserved 384 KB. It was not envisioned that the 640/384 KB limits would be broken under PC and DOS technology.

Later, the Color Graphics Adapter (CGA) appeared. This board introduced PC users to the wonderful world of color and bitmapped ("dot-by-dot") graphics. Video RAM requirements suddenly quadrupled to 64 KB. Because the CGA's RAM address space

Figure 1



began at address "A000 hexadecimal" instead of B000 hex like the MDA, DOS memory was truly cemented to a 640 KB maximum.[1]

Software applications emerged that took advantage of the new color graphics display capabilities (Lotus 1-2-3 and others) and thus needed more RAM in which to operate. By then, IBM had increased the standard complement of PC memory to 256 KB. Lotus 1-2-3 had another peculiarity in that it required a user's entire spreadsheet to reside in RAM. People with very large spreadsheets sometimes ran out of even the full 640 KB of RAM available to DOS. Other applications could read user data or portions of program code from disk during their operation (many still do). This caused the application to run slowly, since the disk would have to be accessed often.

The first memory trick: Expanded Memory

Eventually, "Power users" began clamoring for more usable memory for applications. In response, Lotus, Intel and Microsoft devised a "bank-switched" memory scheme in the mid-1980s called the "Expanded Memory Specification" (EMS). EMS is a "pool" of memory controlled by an "expanded memory manager." The manager can consist either of hardware: an EMS memory board or special circuitry on the 8086/88 or 80286 CPUs, or software: a control program (on the 80386 and above). The actual physical memory ("chips") used by EMS may be located on an EMS board or may be "extended" memory above 1 MB on an 80286 (with EMS circuitry) or above.

[1]The Intel 8088's 1 MB address space is labeled from 000h to FFFh. Each "segment," e.g. 000-0FFh, is 64 KB in length. When using an MDA, DOS memory (which begins at address 0000h) could span to B000h, or 704 KB. But since the original version of DOS required all of its memory to be contiguous (adjacent), the appearance of the CGA confined DOS memory to the space between 0000 and A000h, or 640 KB.

EMS works by "mapping" pages of EMS memory into conventional memory (the 0-1024 KB addressable by the Intel 8088). Mapping means that EMS changes addresses of physical memory located elsewhere, making the memory appear to reside in the conventional memory space. When implemented, EMS creates a 64 KB window called an "EMS page frame" in an unused memory location within the 384 KB space reserved for hardware adapters. Applications make "calls" to EMS, which maps various 16 KB EMS pages to conventional memory via the page frame as requested. The application uses the EMS page frame as a window through which it can read and write 64 KB of data at a time, later changing its "view" to another 64 KB of EMS memory (Figure 2). Up to 32 MB of RAM may be used as EMS memory if it is physically available. The process is very fast because EMS only manipulates the addresses or places to look to find the data; data is not actually copied from one memory location to another.

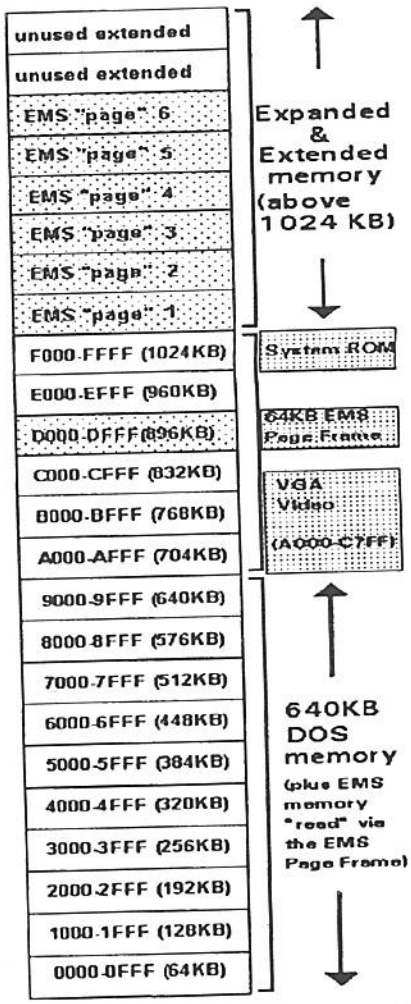
Many applications will use EMS memory. Lotus 1-2-3, for example, uses EMS to store spreadsheet data, while WordPerfect can use EMS to store its large program overlay called WP.FIL (which otherwise is read from disk when needed). Current operating system versions (MS-DOS 5.0) and "operating environments" (DESQview) also use expanded memory to store parts of their code and data, or use EMS to load device drivers and TSRs (terminate and stay resident or "popup" programs). EMS memory can be implemented on any PC.

The second (and third) trick: HIMEM.SYS and extended memory

With the release of DOS 3.x, Microsoft developed a method to use extended memory directly, without converting it to expanded (EMS) memory. This method is known as the "Extended Memory Specification" (XMS). XMS memory is implemented through a device driver called HIMEM.SYS (other drivers also exist). HIMEM.SYS may only be used on PCs that have an 80286 CPU or better (only such PCs can address extended memory). HIMEM provides two features, the *HMA* (High Memory Area) and *XMS memory* Extended Memory Specification memory).

To explain XMS, it is necessary to discuss the difference between the *real* and *protected* modes of the Intel CPU family. The original Intel-1 8088 and 8086 CPU functioned only in *real mode*, in which the CPU can address only 1024 KB of RAM. 80286, 80386,

Figure 2



and 80486 CPUs can operate in either real mode or in *protected mode*. Standard DOS applications use *real mode*; they can only address and manipulate data contained in the 1024 KB of RAM addressable by the Intel 8088. (A real mode application that accesses data stored above 1024 KB only does so because EMS fools it into finding the data within the EMS page frame, below the 1024 KB boundary.) *Protected mode* allows the CPU to access extended memory (above 1024 KB). The word "protected" is used because of additional abilities that allow an operating system to prohibit "bad" applications from crashing the system, and protecting applications from one another's memory space. Programs such as Windows 3.1 and DESQview 386 operate the 80386 CPU in protected mode. Windows uses HIMEM.SYS; DESQview 386 uses the QEMM.SYS expanded/extended memory manager. (QEMM.SYS can be used with Windows also.)

When HIMEM is loaded, it creates a 64 KB area starting at 1024 KB and ending at 1088 KB called the "HMA" (High Memory Area). The HMA can be accessed by the CPU in real mode, thanks to a loophole in the Intel 8088 CPU's design. DOS 5.0 has taken advantage of this loophole by using the HMA to store about 38 KB of DOS program code (the COMMAND.COM command interpreter), freeing up 38 KB of "real" memory for use by standard DOS applications. Other programs, such as DESQview (used to multitask DOS applications), can also use the HMA to store parts of their program code in order to save "real" memory for applications (Figure 3).

HIMEM's other function is to provide extended memory to applications that are written to comply with XMS. HIMEM copies data between extended and conventional memory without the application having to do so itself. It does this by switching the CPU between real and protected modes — real mode to pass data to the application running in DOS conventional memory, and protected mode to access more data that is residing in extended memory. 80286 CPUs may access up to 16 MB of memory using XMS (16,777,216 bytes); 80386s may access up to 4 gigabytes (GB), or 4,294,967,296 bytes.

Other schemes

Just to confuse the issues a bit more, some DOS applications can run in more than the standard DOS memory/real mode limits of 640 KB/1 MB, and rely on neither EMS supplied by expanded memory hardware or software emulators nor XMS supplied by HIMEM.SYS. These applications (such as AutoCAD release 11, Waterloo Maple V, and Lotus 1-2-3 release 3) rely on memory control programs known collectively as **DOS extenders.** DOS Extenders are special utilities or "control programs" that are used with protected mode applications like those mentioned above. When the application makes a real mode request to DOS, the DOS Extender copies necessary data into conventional memory, switches the CPU into real mode, requests the real mode service, and switches the CPU back into protected mode, returning the result of the real mode request into the protected mode application. Phar Lap is the leading producer of DOS Extender software.

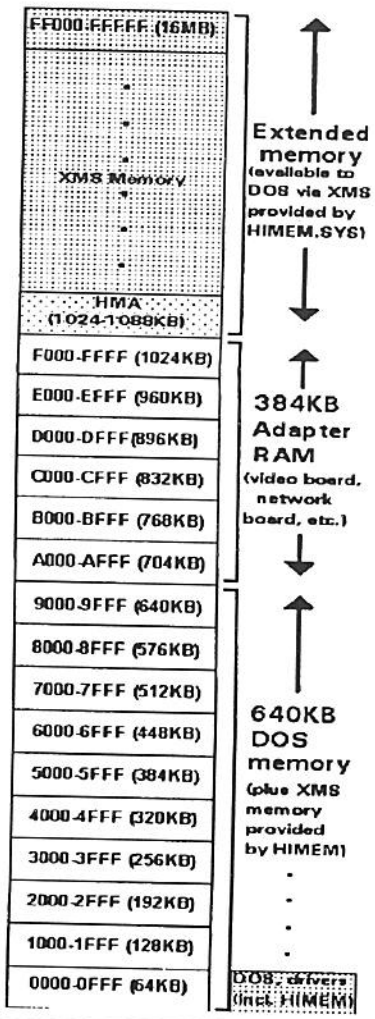
Protected mode applications that use DOS Extenders may or may not be compatible with EMS or XMS memory managers or protected mode control programs such as DESQview and Windows, although most application manufacturers have updated or are updating these programs to ensure that they can be used with the other memory management schemes.

To summarize

PCs running DOS have been increasingly plagued by the problem of "RAM cram": not having enough usable memory to run DOS applications efficiently or at all. The root of this problem lies in the design of the original PC CPU; the Intel 8088 could only manage 1024 KB of RAM. 640 KB of this RAM was given over to use by DOS and applications; the other 384 KB of RAM was allocated to PC hardware adapters.

Over the years, software developers have come up with schemes that fool a DOS PC into using more RAM than it "should." **EMS** (Expanded Memory Specification) may be used by any PC with proper hardware or software; it works by mapping memory normally "out of reach" of the DOS PC into a page frame that can be accessed by DOS applications that support EMS

Figure 3



(many do). **XMS** (Extended Memory Specification) was developed later and can only be used by PCs with 80286 or higher CPUs. It uses extended memory directly without converting it to expanded memory, by switching the CPU between real and protected modes as conventional or extended memory is read or written. A third class of applications run in protected mode only; **DOS Extender** software provides the interface between real mode (for requests to DOS) and protected mode (for most of the application's activity).

Thanks, now what?

Giving memory management advice to people who don't have a trust fund can be dangerous, especially if they are saddled with outdated PCs. So I'll dispense my advice in small doses; take as needed.

Hardware

Information about the capabilities and necessary hardware for various systems is shown in the chart on page 25.

Software

- **EMS Applications** — An application's documentation should clearly state that it will take advantage of EMS memory found in the system. Some programs will use it for data, some for program code, some both, and some none. Most popular up-to-date applications will make some use of EMS. Other possible uses of EMS memory are discussed below.
- **XMS Applications** — Some newer DOS applications are capable of using XMS memory. The most notable (although I call it an "operating environment" rather than an application) is Microsoft Windows 3.x, which uses XMS to multitask Windows applications (e.g. Word for Windows, Excel) on an 80286 or higher with sufficient extended memory. Windows relies on the HIMEM.SYS driver to provide its XMS memory (you can use QEMM.SYS also). XMS memory can be used for other purposes as discussed below.

DOS 5.0

DOS 5.0 comes with both XMS and EMS memory managers. HIMEM.SYS is used to provide the HMA and XMS memory on 80286s and above. DOS 5.0 can use the HMA to store part of COMMAND.COM and DOS disk "buffers," freeing about 38 KB of memory for DOS applications. To do this, the CONFIG.SYS file must contain the lines **device=himem.sys** and **dos=high**. If your PC has at least an 80386SX CPU, you may also use EMM386.EXE, the EMS emulator and manager to convert extended memory into expanded memory and/or "upper memory blocks" (UMBs). Expanded memory can be used to give more usable RAM to DOS applications. UMBs are placed in areas of PC adapter RAM space that are not occupied by adapters. Device drivers (except HIMEM.SYS) and TSR programs can be loaded into UMBs instead of DOS (640 KB) memory, freeing additional DOS memory for applications. A memory configuration like that shown in Figure 4 could be realized by placing the following lines in CONFIG.SYS:

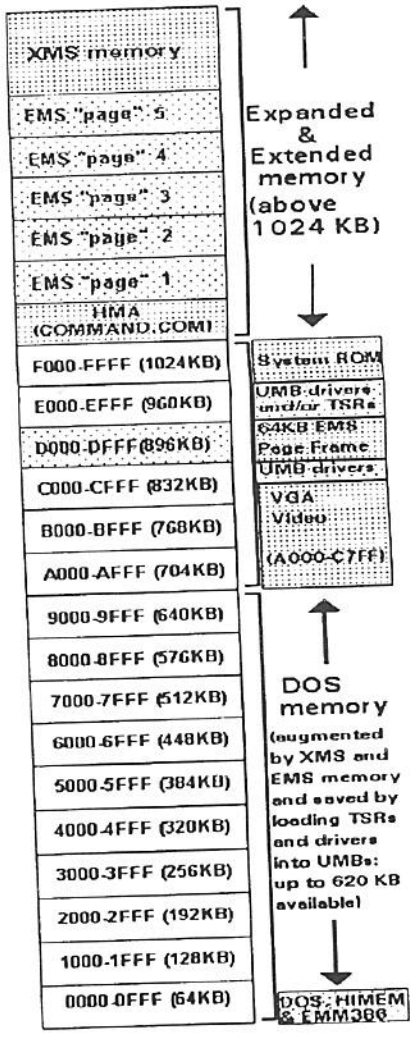
```
device=himem.sys
device=emm386.exe ram 320
frame=d000,dos=high, umb.
```

Device drivers and resident programs may then be loaded into the UMBs with DOS "loadhigh" and "devicehigh" statements, such as **devicehigh=mouse.sys** in CONFIG.SYS or **loadhigh doskey.com** in AUTOEXEC.BAT.

Windows and multitasking

Windows 3.1 requires XMS memory in which to run; HIMEM.SYS (or QEMM) must be loaded (version 3.0 runs with 640 KB on 0886/88s). True Windows applications require only XMS memory, and can be multitasked. To run DOS applications requiring EMS within Windows on 80286 PCs, your PCs expanded memory hardware, driver, and manager must be loaded. On 80386s or better, Windows will provide its own EMS for DOS applications if Windows is run in "386 enhanced mode" (win /3). DOS applications can only be multitasked when Windows is running in 386 enhanced mode. DOS multitasking is possible due to the "virtual 8086 mode" of the 386 or higher CPU; the CPU can remap

Figure 4



memory to make it appear that a DOS application residing in extended memory is running in conventional memory. Protected mode programs such as AutoCAD and Lotus 1-2-3 v3.x can be run with Windows 3.1 but not Windows 3.0. Running Windows is a good choice for those who would like to multitask using a graphical interface.

DESQview and multitasking

DESQview uses EMS predominantly. On 286s, DOS applications can be swapped to EMS memory, and multitasked in EEMS memory. The optional QRAM program from Quarterdeck allows you to load TSRs and device drivers in adapter RAM space in the same way DOS 5.0 uses UMBs. On 386s, DESQview 386 comes with the QEMM memory manager. QEMM enables and manages EMS and XMS memory along with UMBs. It also allows protected mode programs to run concurrently with other DOS applications. Like Windows, DESQview also multitasks using virtual 8086 mode on 386s and up. Running DESQview is a good choice for those who would like to multitask DOS character-based applications.

The Future

One point that has not been brought up in all of the discussion above (but should be fairly obvious if you've read this far), is that memory management on DOS PCs is not a trivial issue. Configuring your PCs memory so that **all** of your programs work to their optimum potential is, at best, a trial and error proposition. What may succeed in allowing one application to run optimally may easily cause another application to fail. Diagnosing such problems can be painstaking. (For help with memory management on your PC at UNT, contact Network and Microcomputer Support at 565-2316).

Worrying less about PC memory management has been a long term goal of software developers. All of the methods above are steps along the way. The future (very near) holds great promise.

IBM has begun shipping OS/2 v2.0, which may become the replacement operating system for DOS if it is successful. It is claimed to be a "better DOS than DOS, better Windows than Windows, and better OS/2 than OS/2." OS/2 2.0 can run multiple DOS, Windows, and OS/2 applications concurrently, and provide more memory for each application than DOS or Windows can: up to 64 MB per Windows session, and up to 633 KB "conventional" memory and 512 MB of extended memory per DOS session, with support for 32 Windows sessions (that's 32 **copies** of Windows), and 32 DOS sessions. All of this "virtual" memory capability requires a minimum of only 4 MB of physical RAM in the PC, plus some extra disk space. Priced at $195, OS/2 2.0 includes both DOS 5.0 and Windows for OS/2. This is a new "32 bit" operating system that was written with the memory management capabilities of the 80386 or better CPU in mind.

Following later this year will be Windows NT, Microsoft's 32 bit version of Windows. It promises improvements similar to OS/2 2.0.

For weary memory managers, both products are worth anticipating.■

CPU	EMS?	XMS?	Multitask under DOS?
8086 or 8088	Yes, with EMS board ($50 to $400), plus RAM (~ $40/MB).	No extended memory. Upgrade PC (minimum 80385SX, preferably 80386 or better, under $750).	No[1]. Upgrade PC (minimum 80386SX, preferably 80386 or better, under $750).
80286	Yes, with EMS board or special built-in system circuitry (NEAT chip set), plus RAM.	Yes, with DOS ($50, preferably 5.0) or other XMS driver, plus RAM. No additional hardware required.	Yes, can run multiple Windows (under $80) but not DOS applications concurrently.[1]
80836 or more [2]	Yes, with DOS or other EMS driver, plus RAM.	Yes, with DOS or other XMS driver, plus RAM.	Yes, with Windows, DESQview (under $100) or other software.

[1] Quarterdeck's DESQview will multitask DOS programs on any PC with an "enhanced expanded memory" (EEMS) board, properly configured, but EEMS boards are difficult to find and high priced. A 386 PC is a better value.
[2] 80386s or better have built-in memory mapping capabilities that, to be used, need only the proper readily available software.

Virus Update

Compiled by Claudia Lynch, *Benchmarks* Editor (BITNET: AS04@UNTVM1)

The following information comes from Valert-l@edu.lehigh.cc.ibm1, UNT VAX News, the SIMTEL20 Archives, and Net-Man: The UNT Network Managers newsletter.

IBM and Compatible PCs

- **New Version of F-PROT** — F-PROT 2.04, the latest release of Fridrik Skulason's antiviral software is now available. This version identifies and disinfects a number of new PC viruses that have cropped up. It also scans through files much faster than previous versions.

  UNT has a site license for F-PROT. To acquire the latest release, you can:

  - Anonymous FTP to ftp.unt.edu. GET the file **fp204.exe** in the pub/antivirus/ibm directory.
  - Request the software from Microcomputer Support at 565-2316 if you are faculty or staff.
  - Go to the ACS General Access Lab in ISB 110 and copy F-PROT onto your own diskettes.
- **Hacked version of PKZIP** — The following was posted to SIMTEL20 by Phil Katz, President, PKWARE Inc.

  There is a file being circulated on BBSs called PKZ201.ZIP or PKZ201.EXE or similar, and it claims to be version 2.01 of PKZIP but in fact is a hacked version of PKZIP 1.93 Alpha.

  As of June 16, 1992, the latest version of PKZIP is Version 1.10. If you see the files PKZ201.ZIP or PKZ201.EXE on any BBS or online system, please ask the SysOp of that system to remove the files immediately, and please contact PKWARE to report where the files were seen.

  Of course, because of this, PKWARE will never release a version of PKZIP with the version number 2.01. If you have any informatin about the source of PKZ201.EXE or PKZ201.ZIP, please report it to PKWARE immediately either:

  - by voice: 414-354-8699
  - by BBS: 414-354-8670
  - by VAX: 414-354-8559
  - or by mail: PKWARE Inc., 9025 N. Deerwood Dr., Brown Deer, WI 53223

Macintosh

- **Antiviral Utilities** — Two freeware antiviral utilities for the MAC are available from ftp.unt.edu. Disinfectant 2.8 and Gatekeeper 1.2.4 are stored in the pub/antivirus/mac directory as the files **disinfectant-28.hqx** and **gatekeeper-124.hqx**.
- **New Macintosh Virus Discovered** — Two slightly different strains of a new Macintosh virus were discovered June 30, 1992. They were included with the game application GoMoku, Versions 2.0 and 2.1. The files were posted to the USENET newsgroup com.binaries.mac, and uploaded to various ftp archives, including the one at sumex-aim.stanford.edu. The game was distributed under a falsified name. The name used in the posting, and embedded in the game, is that of a completely uninvolved person. The actual virus author is unknown, and probably used this person's name as a form of harassment.

  Authors of all major Macintosh antiviral tools are planning updates to their tools to locate and/or eliminate this virus. Some of the programs that are being updated are:

  - Disinfectant — Release 2.9 should be available by July 6.
  - Gatekeeper — Release 1.2.6 will be available "soon."
  - Rival — available now.
  - Virex — Release 3.82 available July 6.
  - VirusDetective — Release 5.0.5 available now.
- **MAC user warning** — Neither Apple System 7 nor 7.0.1 should be used UNMODIFIED because they have the "disappearing folders" bug. You should be sure to have the (free) **System Tuner 1.1.1** from Apple installed on these systems; Versions 1.0 and 1.1 of the Tuner are outdated or buggy and should not be used. Tuner 1.1.1 is available from Authorized Apple dealers, from many user groups, from commercial networks, and on several places on the Internet. Also, System 7.0.1 is a later release of System 7. System 7.0.1 also needs the update installed.■

WordPerfect User's Group Summer Schedule

The WordPerfect User's Group is continuing to meet this summer. The schedule is as follows:

- **July 17, 2-3pm**, SCS Lab - Chilton 255, main focus - Documented but Unadvertised Features of WP 5.1 (Advance, Block Protect, Document Summary, and Window)
- **August 21, 2-3pm**, SCS Lab - Chilton 255, main focus - Sort Function: Line, Paragraph and Secondary Merge File. ■

VAXCLUSTER USAGE STATISTICS

May Top Ten Programs: CPU Time Used

Program	Description	CPU Time	Percent of Total
1. GAUSSIAN	Molecular Modelling	3 11:43:43.31	27.2
2. MOPAC	Quantum Mechanics	1 21:18:38.55	14.7
3. DEFRAG	Disk Optimizer	1 17:02:16.28	13.3
4. NEWS	ANU News Utility	1 11:14:42.99	11.4
5. User programs	Compiled Programs	1 06:05:09.23	9.8
6. BACKUP	Disk Backups	0 10:03:20.43	3.3
7. MAIL_SERVER	VMS Mail Server	0 07:06:40:18	2.3
8. DIRECTORY	VMS Utility	0 06:43:39.83	2.2
9. LOGINOUT	User Login	0 05:19:53.32	1.7
10. MAIL	VMS Mail Utility	0 05:14:47.52	1.7
Total		12 20:17:24.30	

May Top Ten Programs: Frequency of Runs

Program	Description	Number of Runs	Percent of Total
1. SET	VMS Utility	94511	20.4
2. LOGINOUT	User login	83983	13.8
3. DIRECTORY	VMS Utility	64003	9.1
4. DELETE	VMS Utility	61102	13.2
5. SENDFILE	VMS Utility	29940	6.5
6. User programs	Compiled Programs	19679	4.2
7. MAIL_SERVER	VMS Mail Server	13749	3.0
8. SYSLOGIN	User Login	12475	2.7
9. MAIL	VMS Mail Utility	9072	2.0
10. SEND	BITNET Message Utility	7801	1.7
Total		463584	

Sol: A UNIX System for Research

By Marc St.-Gil, UNIX Systems Administrator (mstgil@unt.edu)

For many of you on campus who are doing research work, there is a relatively new computing resource available to you in the form of a Solbourne super-minicomputer. This computer runs an operating system that is derived from SunSoft, Inc.'s SunOS that is in very widespread use on campuses and research environments all over the world. Because OS/MP is application binary compatible with SunOS, Sol, our Solbourne, can run almost all of the vast number of UNIX software products for Sun computers with no changes necessary. Another plus of the Solbourne system is that it is a multi-processor OS and has been in use for several years now whereas SunSoft's multi-processor support is in its infancy. Our Solbourne currently has 4 SPARC CPUs giving it some very respectable performance specifications, 114 MIPS (Millions of Instructions Per Second), 26.5 MFLOPS (Millions of FLoating point Operations Per Second) in single precision, and 15.1 MFLOPS in double precision. Sol also has 192 MBytes of RAM with an additional 411 MBytes of disk based Virtual Memory for a total useable working memory of 603 MBytes. For user and system storage, Sol has 5.7 GBytes of disk storage with another 2.6 GBytes of high speed disk on order. Put all of this together and you see that Sol is a very powerful system in terms of hardware and software availability. Of course all of this means nothing if Sol cannot offer you the software to take advantage of all this raw power.

Academic Computing Services has been putting a great deal of effort into making Sol a pleasant platform for our user community as well as offering a fairly wide variety of computing based resources. In addition to specific research software, we have installed a number of user convenience software packages including **nn** for reading Network News, and **elm** for sending and receiving electronic mail (E-mail). In addition to that we also have the entire wuarchive archives mounted across the Internet that appear as local disk partitions. This allows our users to simply use the UNIX copy command, **cp**, to retrieve the files archived there instead of having to use the **ftp** utility. In addition to these conveniences, you might also want to refer to Billy Barron's article in "The Network Connection" on page 12. The following paragraphs describe available research software packages grouped by function or intended user.

For our statistical user community, we have installed SAS version 6.07 for UNIX on Sol with all available modules for use by anyone on the system. To start SAS, use the command **sas**. Out of these modules one of particular

note is SAS's new version of SAS Assist for X-Windows under UNIX. This module provides Sol's users with a windowed interface to the SAS system and make SAS so easy to use that even I was able to perform useful tasks with SAS the first time I ever ran it. This is significant because although I am quite experienced with UNIX and the X Window System, I have no statistical experience at all and had never used SAS before. A Computing Center handout on using SAS/UNIX on the Solbourne is available free of charge in the Computing Center main office (ISB 119). In person SAS consultation for SAS/UNIX 6.07 is available from the Research and Statistical Services support staff in the Computing Center offices. We are also working to provide access to some or all of the SAS Data Sets that are currently available on the Academic Mainframe to users of the Solbourne. Also available to those interested in doing statistics work is the complete IMSL Statistics Library Set for Fortran programmers. To use the routines in this library you must add the key **-limsl** to the end of your **f77** command. These routines are also available from C, but it requires special coding. Contact me in the Computing Center office for help with this. The IMSL Statistics Library contains routines for approaching problems in the following areas:

- ❑ Basic Statistics
  - Frequency Tabulations,
  - Univariate Summary Statistics,
  - Ranks & Order Statistics,
  - Parametric Estimates and Tests,
  - Grouped Data, and
  - Continuous Tabular Data
- ❑ Regression
- ❑ Correlation
- ❑ Analysis of Variance
- ❑ Categorical and Discrete Data Analysis
- ❑ Nonparametric Statistics
- ❑ Tests of Goodness of Fit and Randomness
- ❑ Time Series Analysis and Forecasting
- ❑ Covariance Structures and Factor Analysis
- ❑ Discriminant Analysis
- ❑ Cluster Analysis
- ❑ Sampling
- ❑ Survival Analysis, Life Testing, and Reliability
- ❑ Multidimensional Scaling
- ❑ Density and Hazard Estimation
- ❑ Line Printer Graphics
- ❑ Probability Distribution Functions and Inverses
- ❑ Random Number Generation
- ❑ Utilities
  - Printing,
  - Permutations,
  - Sorting,
  - Searching,
  - Character String Manipulation,
  - Time,
  - Date, and
  - Data Set Retrieval.

Documentation for this software library is available in the Science & Technology Library. It is kept with the "reserve" books under the name Billy Barron and must be requested by the name "IMSL STAT/LIBRARY."

For people doing mathematical research or research that involves com-

SOLBOURNE USAGE STATISTICS

May Top Ten Programs: CPU Time Used

Program	Description	CPU Minutes	Percent of Total
1. g90	Gaussian 90	52819.8	61.4
2. UserPgm	User Programs	23584.1	27.4
3. sas	Statistical Package	4363.1	5.1
4. xterm	X Terminal	1298.4	1.5
5. ircd	Internet relay Chat Daemon	667.5	0.8
6. find	File Locator Utility	587.8	0.7
7. test	Comparison	453.6	0.5
8. snmpget	SNMP Manager	332.4	0.4
9. update	Filesystem sync	280.7	0.3
10. ypserv	Yellow Pages Server	222.7	1.1
Total		86081.9	

May Top Ten Programs: Run Time Used

Program	Description	Number of Runs	Percent of Total
1. snmpget	SNMP Manager	107548	16.0
2. UserPgm	User Programs	83766	12.5
3. sh	Bourne Shell	63498	9.5
4. inetd	Internet Master Daemon	35112	5.2
5. csh	C Shell	25167	3.8
6. sendmai	Mailer Program	20822	3.1
7. ls	Directory	20387	3.0
8. echo	Print	16244	2.4
9. test	Comparison	16032	2.4
10. rm	Delete File	15345	2.3
Total		671013	

plex mathematical manipulations, we have the IMSL Math Library Set for Fortran. To use the routines in this library you must add the key **-limsl** to the end of your **f77** command. These routines are also available from C, but it requires special coding. Contact me in the Computing Center offices for help with this. The IMSL Math Library contains routines for approaching problems in the following topics:

- ❑ Linear Systems
- ❑ Eigensystem Analysis
- ❑ Interpolation and Approximation
- ❑ Integration and Differentiation
- ❑ Differential Equations
- ❑ Transforms
- ❑ Nonlinear Equations
- ❑ Simple Matrix and Vector Operations
- ❑ Various Calculation Optimization Routines

The documentation is in the same place as the STAT/LIBRARY documentation and is called "IMSL MATH/LIBRARY." Also available in the IMSL Math Library is a set of what IMSL calls "Special Functions." See above for instructions on accessing this library. These functions are in the following areas:

- ❑ Elementary Functions
  - Roots,
  - Bases, and
  - Logarithms
- ❑ Trigonometric and Hyperbolic Functions
- ❑ Exponential Integrals
- ❑ Gamma Functions
- ❑ Error Functions
- ❑ Bessel Functions
- ❑ Kelvin Functions
- ❑ Airy Functions
- ❑ Elliptic Integrals
- ❑ Elliptic Functions
- ❑ Probability Distribution Functions and Inverses
- ❑ Mathieu Functions
- ❑ Miscellaneous
  - Spence Dilogarithms
  - Chebyshev Series

Also available for mathematics research is Mathematica for SunOS. This package offers most of the functionality that Mathematica users on PCs, Macs, and NeXTs are familiar with including the ability to display static and animated graphics of functions. The only thing missing is the "Notebook" user interface. A work around for this is available by using the PC, Mac, or NeXT version as a "front end" and letting the Mathematica "kernel" on Sol handle the actual computations. This gives a user all the convenience of the Notebook interface with the superior number crunching ability of the Solbourne CPUs. The Mathematica program is invoked locally with the **math** command. To use your local copy of Mathematica as a front end to Sol's version, see your Mathematica documentation.

For those interested in chemistry there is a software package called Gaussian 90 available through special permission from a Chemistry Department faculty member. Since the interest in the package was strong, but limited to a small group of people in one department, we allowed that department to purchase the software and then installed the software on Sol with special usage restrictions. Those interested in using this software package should contact Academic Computing Services at 565-2324.

For those interested in programming, Sol currently offers C and ANSI FORTRAN 77 compilers from Solbourne invoked as **cc** and **f77** respectively. Sol's FORTRAN also supports most VMS FORTRAN extensions so porting FORTRAN from a VAX running the VMS operating system should be very simple. The MIT Scheme interpreter, **scheme**, is available and we will be installing a Common LISP product in the near future. Also, available are GNU ANSI C and GNU C++ compilers which are invoked as **gcc** and **g++**, respectively. ■

The Coalition for Networked Information

This article appeared in Reach, the newsletter of the Humanities Computing Facility of the University of California at Santa Barbara (Fall 1991, p. 2).

The Association of Research Libraries, CAUSE, and EDUCOM have joined together to form the Coalition for Networked Information.

Its objective, as stated in the announcement of its formation, is "to promote the creation of and access to information resources in networked environments in order to enrich scholarship and to enhance intellectual productivity."

A number of working groups have been formed to examine means of achieving these goals. Among them are groups on Transformation of Scholarly Communication, Modernization of Scholarly Publication, Access to Public Information, Directories and Resource Information Services, and Teaching and Learning.

The Coalition will be collecting and organizing information about experiments and demonstration projects in the application of advanced networks and high performance computers to research and educational communication processes.

It will try to identify and establish suitable means for the further testing and evaluation of such innovative projects in computer use.

Information about the plans and activitries of the Coalition will be distributed through the publications of the Association of Research Libraries, CAUSE, and EDUCOM.

For further information on the Coalition, please communicate with:

Coalition for Networked Information
1527 New Hampshire Ave., N.W.
Washington, DC 20036
Phone: 202-232-2466 ■

Mainframe Performance Statistics

May Top Ten Programs : Frequency Of Runs

Program	Description	# of Runs	% of Total
1. IEBGENER	IBM Utility	6977	55.9
2. SPSS	SPSS Version 4.0	941	7.5
3. SASLPA	SAS Version 5.18	889	7.1
4. PGM=*.DD	Compiled Program	470	3.8
5. IEWL	Linkage Editor	463	3.7
6. IGYCRCTL	VS COBOL2 Compiler	245	2.0
7. IEFBR14	IBM Null Utility	219	1.8
8. IDCAMS	VSAM Utility	163	1.3
9. ST033	UNT Stat. Procedure	147	1.2
10. NOTFACE	UNT Stat. Procedure	145	1.2

May Top Ten Programs: CPU Seconds Used

Program	Description	CPU Seconds	% of Total
1. SASLPA	SAS Version 5.18	28486	37.2
2. PGM=*.DD	Compiled Program	16543	21.6
3. SPSS	SPSS Version 4.0	8974	11.7
4. COMPLET4	Academic COM-PLETE	6913	9.0
5. SAS370	SAS Version 6.06	2980	3.9
6. SSS4001	Operations Automation	2957	3.9
7. ISTINM01	VTAM Utility	2242	2.9
8. IEBGENER	IBM Utility	1423	1.9
9. ADARUN	ADABAS Utility Module	863	1.1
10. RESOLVE	MVS Utility	500	0.7

Operating Systems Performance Statistics for May

CPU	SYSTEM	Planned Production Hours	Production Hours Achieved	System Uptime
ACAD	VM/XA	744.00	742.33	99.8%
ACAD	MUSIC/SP	724.45	722.53	99.7%
ACAD	MVS/JES2	744.00	741.40	99.7%
ACAD	COMPLETA	743.38	740.59	99.6%
ADMN	MVS/JES2	735.05	732.34	99.6%
ADMN	COMPLETA	272.00	272.00	100.0%
ADMN	ADABASA	707.22	707.22	99.6%

- The ACAD CPU (HDS/8083) achieved 100% uptime in May. The HDS/7360 DASD achieved 100% uptime in May. The HDS/7380 DASD achieved 100% uptime in May.

- The ADM CPU (IBM 9121/440 processor) achieved 100% uptime in May. The HDS/7360 DASD achieved 100% uptime in May. The HDS/7380 DASD achieved 100% uptime in May. The IBM 3390 DASD achieved 100% uptime in April. The EMC Solid State Disk achieved 100% uptime in May. ■

Key Causes Of Lost Productivity In May: ACAD CPU

Miscellaneous

1. VM/XA system maintenance. 2.79 HOURS

Key Causes Of Lost Productivity In May: ADMN CPU

Miscellaneous

1. Scheduled microcode upgrade on the IBM 9121 processor. 9.33 HOURS
2. Preparatory systems maintenance for implementation of the IBM MVS/ESA operating system. 2.95

GRAND TOTAL 12.28 ■

Disk Backup Schedules

SYSTEM	BACKUP	DESCRIPTION
Administrative MVS/SP	Daily	Monday - Friday around 7 p.m. (after COM-PLETE is shut down) & on Saturday & Sunday if COM-PLETE has been up that day.
	Weekly	Full pack dumps taken each Sunday morning.
	Monthly	Full pack dumps taken on the first day of each month.
Academic MVS/SP	Daily	Monday - Sunday during the early hours of the morning.
	Weekly	Full pack dumps taken each Sunday.
	Monthly	Full volume dumps taken on the first day of each month.
MUSIC/SP	Daily	Wednesday - Monday starting at 4 a.m. and lasting about 30 minutes.
	Weekly	Tuesday mornings at 3 a.m., these last about 2 hours.
	Semester	Once a semester, a permanent backup is taken.
VM/XA	VM Weekly	Early every Wednesday morning.
	CMS mini-disks	Daily backup performed early every morning. Weekly backup every Tuesday starting after Midnight.
	Semester	Once a semester, a permanent backup is taken.
VAXcluster	Daily	Incremental backups are performed Monday - Thursday at 6 p.m. Saturday & Sunday at 5 p.m.
	Weekly	Full backups are performed every Friday beginning at 8 a.m. Generally lasts all day.
	Monthly	A "stand-alone" backup is performed monthly. Dates and times are given in the system log-on message.
	Semester	Once a semester, a permanent backup is taken.
Solbourne	Daily	Incremental backups are performed Sunday - Friday at 2 a.m.
	Weekly	Full backups are performed every Saturday at 7 a.m.
	Semester	Once a semester, a permanent backup is taken.

University of North Texas Computing Center

Dr. Philip Baczewski, Assistant Director of Academic Computing, *Benchmarks* Associate Editor
Bill Buntain, Director of Network and Microcomputer Services
Richard A. Harris, Associate Vice President for Computing
Coy Hoggard, Senior Director of Administrative Computing
Steve Minnis, Director of Mainframe Technical Services
Dr. Paul Gandel, Senior Director of Academic Computing
Cynthia Koepp, *Benchmarks* Assistant Editor
Claudia Lynch, *Benchmarks* Editor

Get a Subscription to *Benchmarks*

Benchmarks is a vital link between the UNT Computing Center and the users of our facilities. It is important for all users of the computing facilities to maintain a file of these newsletters because they contain materials which will periodically update existing documents as well as information and suggestions on uses of OS/MVS, MUSIC/SP, CMS, the VAXcluster, Solbourne, Microcomputers, and other resources available to UNT students and faculty. To facilitate the dispersal of *Benchmarks*, *** **FREE** *** subscriptions are available. To receive yours, send the following information to us either by snail mail (the post office or campus mail), FAX (817) 565-4060, or through electronic mail, to the User-ID AS04 on MUSIC, VMS, SOL, or CMS.

Name: _____

Mailing Address: _____

_____

_____

Academic Computing Services
The Computing Center
NT Box 13495
University of North Texas
Denton, TX 76203
FAX 817-565-4060